DOCUMENT RESUME

ED 195 276 IR 009 055

AUTHOR	Smalley, Donald A.: And Others
TITLE	Linking the Bibliographic Utilities: Benefits and Costs. Technical Report.
INSTITUTION	Battelle Memorial Inst., Columbus, Ohio. Columbus Labs.
SPONS AGENCY	Council on Library Resources, Inc., Washington, D.C.
PUB DATE	15 Sep 80
NOTE	314p.: For related document, see IR 009 053.
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	Academic Libraries: Cataloging: Cost Effectiveness: *Data Bases: *Information Networks: Interlibrary Loans: *Library Services: Methods: Models: *Online Systems: Public Libraries: Reference Services: *Shared Services
IDENTIFIERS	*Bibliographic Utilities: Linkage Analysis

ABSTRACT

This study of the benefits, methods, and costs of linking two or more bibliographic utilities for use by public and academic libraries assessed the impact of linking on three distinct areas of library operations: shared cataloging, interlibrary loan, and reference searching. The analysis was restricted to bibliographic records for monographs in the OCLC, RLG/RLIN, WLN, and Library of Congress databases. The findings and analyses of data are presented in two parts. The effects of linking on libraries and their patrons are discussed, and detailed analyses of the benefits of linking are provided for each of the three areas studied in the first part. The second focuses on the technical requirements for establishing a link and three promising alternatives are identified: tape exchange of requested records, online access using the "native mode" of the linked utility, and online access with automatic translation of searches and responses. Requirements for developing each of these alternatives are described and comparisons are made. A summary of conclusions and recommendations concludes the report. Appendices include supporting data, a glossary, more detailed discussions of several topics, and a summary description of the BIBLINK model. A bibliography and several reference lists are provided. (RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195276

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TECHNICAL REPORT

on

LINKING THE BIBLIOGRAPHIC UTILITIES: BENEFITS AND COSTS

Submitted to

THE COUNCIL ON LIBRARY RESOURCES

September 15, 1980

by

Donald A. Smalley
William G. Griffith
Ann M. Walker
Michael B. Wessells

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Lee Jones

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009055

ACKNOWLEDGEMENTS

Battelle gratefully acknowledges the contributions of many people to this study. Members of the administrative and technical staffs at OCLC, RLG, WLN, and the Library of Congress provided much information about their operations, often taking time from their hectic schedules to track down elusive details. Many individual libraries contributed materials for the test searches, and many others completed questionnaires or participated in telephone interviews. In addition, the project team discussed linking with many people in libraries, state and regional networks, and other organizations throughout the country. We take this opportunity to thank everyone for their interest and participation and to express our sincere appreciation to the Council on Library Resources for its assistance and encouragement throughout this study.

The BIBLINK computer model was programmed by Lawrence L. Rose in Battelle's NUCLEUS modeling language. Other Battelle staff members who participated in the study are Allan Bowers, Bruce Buxton, John Fried, George Juras, Gabor Kovacs, Steven Miller, Robert Niehoff, and Janice Patton. A. Hank Epstein of Information Transform Industries and Susan Miller of the Ohio State University Libraries were consultants to the project.

TABLE OF CONTENTS

Chapter		Page
1	INTRODUCTION	1
	Scope of the Study	1
	Background	3
	Organization of the Report	11
2	THE BENEFITS OF LINKING FOR SHARED CATALOGING	13
	Introduction	13
	Methodology	15
	Results of the Current Title Test Search	19
	Current Title Hit Rates	19
	Incremental Hit Rates	26
	The Economic Benefits of Linking for Shared Cataloging	33
	A Model for Calculating Cataloging Cost Avoided	38
	The Cost of Linking for Shared Cataloging	42
	Other Possible Benefits from Linking	44
	Linking for Authority Data	45
3	THE BENEFITS OF LINKING FOR INTERLIBRARY LOAN	47
	Introduction	47
	Verification Benefits	48
	Location Benefits	49
	Methodology	52
	Results of the Borrowing Request Test Search	54
	Verification Hit Rates	54
	Incremental Verification Hit Rates	56
	More Flexible Online Searching	58
	Holdings Hit Rates	59
	Incremental Holdings Hit Rates	62
	The Economic Benefits of Linking for ILL	67
	A Model for Calculating Borrowing Cost Avoided	67
	The Costs of Linking for ILL	74
	Summary	75

TABLE OF CONTENTS (Continued)

Chapter Page

4 THE BENEFITS OF LINKING FOR REFERENCE SEARCHING . . . 77

Introduction 77
Resources Required to Replace Local Catalogs with Online Union Catalogs 77
The Benefit of Linked Reference Searching. . . . 79

A Model for Estimating the Volume of Linked Reference Searching. 81
Local Catalog Traffic 84
Reference Searches of the Host Utility's Data Base 86
The Volume of Linked Reference Searching. . 88
Successful Linked Searches. 91
Conclusions 92

5 DERIVATION AND PRELIMINARY ANALYSIS OF LINK ALTERNATIVES. 95

Introduction 95
What is a Link Alternative?. 97

Exchange Method Options 102
Data Exchanged Options. 103
Communication Method Options. 106
User Interface Options. 109

The Final Link Alternatives. 110

Direct Tape Delivery via Mail of Records Requested in Native Mode. 113
Online/Native Mode via Dedicated Line . . . 115
Online/Translation Mode via Dedicated Line. 117

6 A COMPARATIVE EVALUATION OF THREE LINK ALTERNATIVES . 121

Introduction 121
The Benefits to Libraries of Each Alternative. . 122

Timeliness of Access. 123
Ease of Use 125

TABLE OF CONTENTS
(Continued)

Chapter		Page
6 (Cont.)	Computer Resource Requirements for Each Alternative	126
	Anticipated Search Loads from Linking	127
	Daily Computer Processing Requirements	132
	Incremental Costs of Computer Processing and Data Transmission	134
	Development Effort	139
	Staff Time Requirements	139
	Staff Cost Estimates	145
	Summary of the Evaluation	147
	Implementation Schedule	149
7	CONCLUSIONS AND RECOMMENDATIONS	153
	Summary of Major Conclusions	153
	The Benefits of Linking for Shared Cataloging	153
	The Benefits of Linking for Interlibrary Loan	154
	The Benefits of Linking for Reference Searching	156
	The Method of Linking	158
	The Impact on the Utilities	159
	Recommendations	161
	Continue Analysis and Planning Using the BIBLINK Computer Model	161
	Develop an Online Link with Automatic Translation of Requests and Responses	161
	Appoint Committees to Establish Standards	162
	BIBLIOGRAPHY	163

TABLE OF CONTENTS
(Continued)

		Page
APPENDIX 1	GLOSSARY	A1-1
APPENDIX 2A	THE IMPACT OF LINKING ON THE TIMING OF CURRENT TITLE CATALOGING	A2-1
APPENDIX 2B	COMPOSITION OF THE SAMPLE USED IN THE CURRENT TITLE TEST SEARCH	A2-4
APPENDIX 2C	WORKSHEET FOR CALCULATING ANNUAL CATALOGING COSTS AVOIDED BY LINKING	A2-5
APPENDIX 2D	AUTOMATED AUTHORITY FILES	A2-6
APPENDIX 3A	THE EFFECTS OF UTILITY USE ON ILL COSTS	A3-1
APPENDIX 3B	ILL REQUEST PATTERNS	A3-2
APPENDIX 3C	COMPOSITION OF THE SAMPLE USED IN THE MONOGRAPH BORROWING REQUEST TEST SEARCH	A3-7
APPENDIX 3D	ANALYSIS OF REQUESTS FOR WHICH NO HOLDINGS WERE FOUND	A3-8
APPENDIX 3E	WORKSHEET FOR CALCULATING THE VERIFICATION COSTS AVOIDED BY LINKING	A3-9
APPENDIX 3F	REVIEW OF LITERATURE ON THE COST OF ILL	A3-11
APPENDIX 4A	ONLINE CATALOG REQUIREMENTS	A4-1
APPENDIX 4B	WORKSHEET FOR ESTIMATING THE VOLUME OF LINKED REFERENCE SEARCHING	A4-51
APPENDIX 5A	COMPUTER PROGRAMS REQUIRED FOR THE TAPE DELIVERY/ NATIVE MODE LINK ALTERNATIVE	A5-1
APPENDIX 5B	COMPUTER PROGRAMS REQUIRED FOR THE ONLINE NATIVE MODE LINK ALTERNATIVE	A5-2
APPENDIX 5C	COMPUTER PROGRAMS REQUIRED FOR THE ONLINE/ TRANSLATION MODE LINK ALTERNATIVE	A5-3
APPENDIX 6A	SUMMARY DESCRIPTION OF BIBLINK	A6-1

TABLE OF CONTENTS
(Continued)

Page

APPENDIX 6B LINKED SEARCH LOAD ESTIMATES. A6-13

APPENDIX 6C OVERVIEW OF THE BIBLIOGRAPHIC UTILITIES A6-19

APPENDIX 6D IMPLEMENTATION PLAN FOR THE TAPE DELIVERY/NATIVE MODE LINK ALTERNATIVE A6-20

APPENDIX 6E IMPLEMENTATION PLAN FOR THE ONLINE/NATIVE MODE LINK ALTERNATIVE A6-22

APPENDIX 6F IMPLEMENTATION PLAN FOR THE ONLINE/TRANSLATION MODE LINK ALTERNATIVE. A6-25

LIST OF FIGURES

Page

Figure 1-1. Resources and Services Potentially Available Through Linking 6

Figure 2-1. Hit Rates for Current Titles. 20

Figure 2-2. Graph of Hit Rates for Current Titles 20

Figure 2-3. Overlap of Data Base Coverage for Current Titles. . . 21

Figure 2-4. Hit Rates on LC-MARC Records. 22

Figure 2-5. Hit Rates for Current Titles--Titles Categorized by Type of Contributing Library. 24

Figure 2-6. Hit Rates for Current Titles--Titles Categorized by Utility of Contributing Library 24

Figure 2-7. Hit Rates for Current Titles--Titles Categorized by Type of Material. 25

Figure 2-8. Hit Rates for All Current Titles on Combinations of Data Bases. 27

Figure 2-9. Week 8 Hit Rates for Current Titles on Combinations of Data Bases 28

Figure 2-10. Incremental Hit Rates for All Current Titles. 29

Figure 2-11. Incremental Hit Rates for All Current Titles When All Data Bases are Linked. 29

Figure 2-12. Original Cataloging Avoided by Linking. 31

Figure 2-13. Incremental Hit Rates for Current Titles Contributed by Libraries From Each Utility. 34

Figure 2-14. Incremental Hit Rates for Current Titles Contributed by Three Types of Libraries 35

Figure 2-15. Incremental Hit Rates for Current Non-English Titles. 36

Figure 2-16. General Model for Calculating Annual Cataloging Costs Avoided by Linking. 37

Figure 2-17. Calculation of the Cost Per Linked Search That Would Offset the Cost Avoided 43

LIST OF FIGURES
(Continued)

Page

Figure 3-1 Verification Hit Rates 55

Figure 3-2 Comparison of Hit Rates for ILL Verification and Current Title Cataloging 55

Figure 3-3 Incremental Verification Hit Rates for All Titles. . . 56

Figure 3-4 Holdings Hit Rates on Combinations of Data Bases . . . 60

Figure 3-5 Incremental Holdings Hit Rates 63

Figure 3-6 Incremental Holdings Hit Rates for Requests Contributed by Libraries from Each Utility 64

Figure 3-7 Incremental Holdings Hit Rates for Requests Contributed by Three Types of Libraries 65

Figure 3-8 Incremental Holdings Hit Rates for Non-English Monographs 66

Figure 3-9 Model for Calculating the Verification Costs Avoided by Linking 68

Figure 3-10 Model for Calculating the Location Costs Avoided by Linking. 69

Figure 3-11 Summary of ILL Costs Avoided by Linking. 73

Figure 4-1 Model for Estimating the Benefit of Linked Reference Searching. 82

Figure 4-2 Derivation of Total Linked Reference Searches per Day. 90

Figure 5-1 Network Configuration for Linking. 96

Figure 5-2 Options Within Each Characteristic of a Link Alternative. 100

Figure 5-3 Usage of Member-Generated Records at OCLC. 104

Figure 5-4 Options Remaining After Eliminations 110

Figure 5-5 Direct Tape Delivery via Mail of Records Requested in Native Mode 114

LIST OF FIGURES (Continued)

Page

Figure 5-6. Online Access in Native Mode via Dedicated Line . . . 116

Figure 5-7. Online Access in Translation Mode via Dedicated Line. 118

Figure 6-1. Benefits to Libraries from Each Link Alternative. . . 126

Figure 6-2. Estimated Number of Search Requests Received Each Day From Users of Other Utilities 128

Figure 6-3. Estimated Number of Search Requests Received Each Day From Users of Other Utilities 129

Figure 6-4. Estimated Number of Search Requests Received Each Day From Users of Other Utilities 130

Figure 6-5. Computer-Hours Per Day Required to Process Search Loads Due to Linking. 133

Figure 6-6. Incremental Annual Costs Due to Linking, 1980 Estimates 135

Figure 6-7. Incremental Annual Costs Due to Linking, 1983 Estimates 136

Figure 6-8. Estimated Development Effort for Each Link Alternative 140

Figure 6-9. Estimated Development Effort if Common Communications Processor is Used 144

Figure 6-10. Projected Salary Expense for Development of Each Link Alternative. 146

Figure 6-11. Summary of Comparative Evaluation of Three Link Alternatives. 148

Figure 6-12. Implementation Schedule Showing Overlap of Development Activities. 150

CHAPTER 1

INTRODUCTION

Scope of the Study

This report presents the findings of a study that examined the methods, benefits, and costs of linking the "bibliographic utilities". The study focused on the following selected issues within this broad and complex topic:

(1) OCLC, RLG/RLIN, WLN, and the Library of Congress were considered to be the initial candidates for linking. These organizations have been among the first to consider linking.

(2) The study attempted to determine whether public and academic libraries would benefit from linking two or more of the utilities. Corporate and other special libraries were not included.

(3) The analysis was restricted to bibliographic records for monographs, because the data bases maintained by the utilities and LC cover principally monographic material.

NOTE: Appendix 1 contains a glossary of selected terms used in this report. Whenever appropriate, the definitions given in the NDO Network Planning Paper No. 2 are adopted here.

(4) The study assessed the impact of linking on three distinct areas of library operations: shared cataloging, interlibrary loan, and reference searching. There may also be benefits for other processes and services such as acquisitions and collection development, but these areas were not part of the study. Linked access to OCLC's ILL Subsystem was not studied, and the report deals instead with linking to aid in verification and location.

(5) The Library of Congress is a significant source of online bibliographic data that could be useful to many libraries. The study closely examined only the benefits to the cataloging function, although reference searching and ILL verification are potentially important uses of a link to the LC data base as records not distributed via the MARC subscription service become available at LC. The Library of Congress would also be an important user of other data bases, and the study projected the benefits that LC catalogers could derive from linking.

(6) The technical evaluation concentrated on link alternatives that are technologically feasible today and that could be implemented within the next 3 or 4 years.

Background

The bibliographic utilities have come about in a variety of ways over the past 10 years. This section reviews that evolution briefly and looks at the emerging interest in linking the utilities into a nationwide bibliographic network.

The concept of computer-based resource sharing among libraries began to take shape in the early 1960's, notably in the work of the National Library of Medicine to establish the MEDLARS I network. In another early effort, the medical libraries at Harvard, Yale, and Columbia devised a plan to share their cataloging by exchanging records in a common machine-readable format. The first large-scale application of computers to resource sharing came in 1968, however, when the Ohio College Library Center inaugurated its service to 47 charter member libraries in that state.

Since OCLC began its online service in 1971 and expanded beyond Ohio academic libraries, its union catalog has grown to over 6 million bibliographic records and more than 70 million holdings. The number of services provided to its more than 2,100 affiliated libraries has increased in the past several years, and new services are being developed. This rapid growth of OCLC reflects the desire of libraries to control bibliographic processing costs and the ability of a computer-assisted resource-sharing system to slow the rate of rise of those costs.

The Research Libraries Group (RLG) was formed in 1974 to meet the unique needs of research libraries through a number of ongoing programs. In 1978, RLG inaugurated one of these programs, the Research Libraries

Information Network (RLIN), using the BALLOTS system that had been developed at Stanford University. RLG membership has expanded swiftly in recent months and now totals 20 research libraries. About 50 other institutions which had been participants in the BALLOTS network prior to the formation of RLG continue to use the RLIN system today.

In 1976, the Washington state legislature authorized funding for the Washington Library Network (WLN), a bibliographic utility begun a few years earlier to facilitate resource-sharing among the libraries in that state. With this support, WLN has grown to include about 60 institutions in the Pacific Northwest. The organization became self-supporting in 1979. WLN differs from OCLC and RLG in that it is a regional service and does not intend to expand direct WLN services into other areas of the country.

The bibliographic utility concept has also taken hold in Canada, where the University of Toronto Library Automated Systems (UTLAS) provides bibliographic services to about 500 libraries.

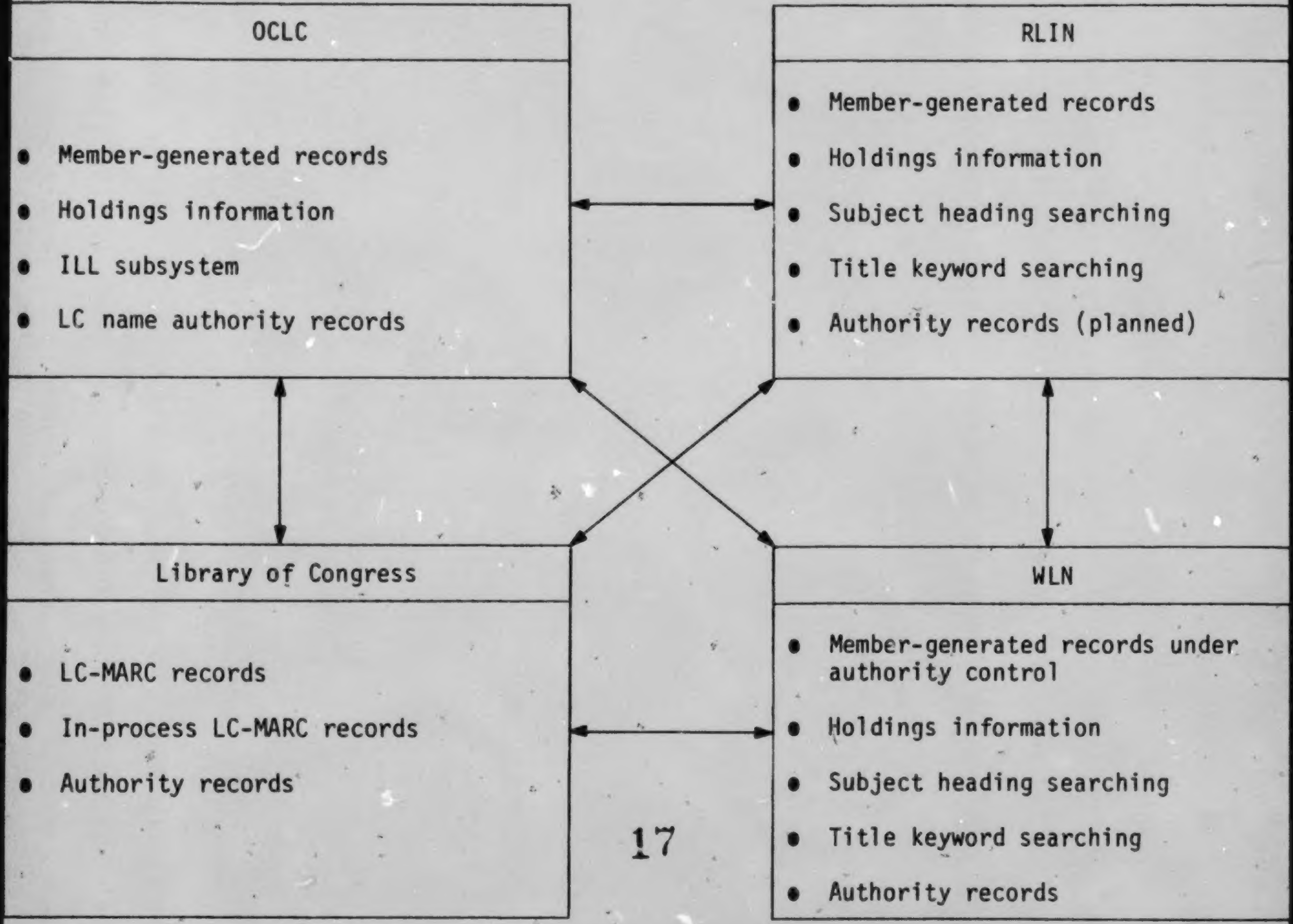
A library that participates in any of these networks avoids some of the costs that it would otherwise incur for technical services like cataloging, card production, and acquisitions. In addition, access to the online union catalog maintained by each utility enhances a library's ability to assist patrons whose needs cannot be met by the library's own resources. As each utility continues to develop, and as more libraries perform retrospective conversions of their bibliographic records, utility users and their patrons are likely to reap still greater benefits.

Most libraries do not have access to more than one utility, and this fact, coupled with the belief that there is much unique information in each data base, has led to recommendations that the utilities consider linking their data bases. Some of the interest in linking the bibliographic utilities can also be attributed to the availability of different services from each utility. If libraries can benefit from participation in a single bibliographic network, the reasoning goes, then they might benefit even more if they had access to resources and capabilities that are not available from their "host" utility.

The Library of Congress, though not technically a utility, is the largest single producer of original cataloging records in machine-readable form. Through its MARC distribution service, LC has stimulated a rudimentary form of linking--the replication of LC-MARC records at each utility. Thus, users of OCLC, RLIN, and WLN have access to bibliographic records that are produced by the Library of Congress. At LC, the cataloging data is maintained on an online computer system for internal use. LC has explored and tested the possibility of offering online access to its bibliographic files to outside parties and, although there is no activity in this area now, LC will likely play an important role in any linking project.

Figure 1-1 summarizes the important features of each utility and the Library of Congress, showing the resources and capabilities at each that might be useful to libraries if they could select services from any of the organizations. For example, access to OCLC's holdings information might help WLN and RLIN users locate additional items to fill the interlibrary loan needs of their patrons. OCLC users might in turn benefit from the use of the subject searching capability at WLN and RLIN.

FIGURE 1-1. RESOURCES AND SERVICES POTENTIALLY AVAILABLE THROUGH LINKING


OCLC
• Member-generated records
• Holdings information
• ILL subsystem
• LC name authority records
RLIN
• Member-generated records
• Holdings information
• Subject heading searching
• Title keyword searching
• Authority records (planned)
Library of Congress
• LC-MARC records
• In-process LC-MARC records
• Authority records
WLN
• Member-generated records under authority control
• Holdings information
• Subject heading searching
• Title keyword searching
• Authority records

The ideal result of linking the bibliographic utilities is to give library users access to a large proportion of the machine-readable bibliographic records in the nation, and perhaps to form one element of an eventual international bibliographic network. Bibliographic records are produced by a wide range of sources, many of which might one day be linked in some fashion. For example, the National Library of Medicine maintains 18 data bases, including bibliographic and non-bibliographic data in bio-medical areas. Online access to some of these data bases is available from NLM and through commercial brokers. The National Agricultural Library (NAL) is pursuing similar developments. The National Library of Canada is another major producer of high-quality machine-readable bibliographic records. Several large libraries have also developed the capacity to produce machine-readable bibliographic records and might serve as additional sources of bibliographic data. For example, the collections at the University of Chicago and the University of Georgia might be made available to other libraries if their data bases were linked with one or more bibliographic utilities and conformed to existing standards.

The subject of linking, interpreted broadly, is concerned with how a nationwide bibliographic network might come into being and what the effects would be on all parties--patrons, libraries, regional networks, and utilities. It has been assumed by many, however, that a logical first step toward creating a nationwide network would be to link the bibliographic utilities. Such a link would make a substantial portion of the nation's machine-readable bibliographic data available to libraries using the services of a single utility.

Many individuals from the utilities, the Library of Congress, associations and funding agencies have explored the idea of linking for several years. LC's Network Advisory Committee (NAC), through its Network Technical Architecture Group (NTAG), did some preliminary work on the technical requirements for a nationwide network. A task force established by the National Commission on Libraries and Information Science (NCLIS) and the National Bureau of Standards (NBS) has developed a computer communications protocol for use in linking bibliographic systems.

Prior to the establishment of RLIN, RLG participated in a pilot test involving the linking of the New York Public Library (NYPL) computer to the Library of Congress' computer system for online retrieval of LC bibliographic records. The Columbia University Library also participated in this project. Columbia terminals were used to search the NYPL data base, with unsuccessful searches switched automatically to the LC data base. A record located on the LC data base was then transferred to the NYPL computer, modified as required by Columbia, and stored in the NYPL data base.

RLG, WLN, and the Library of Congress have recently begun work on a project to link their authority files. The Linked Authority System Project (LASP) is now designing an online link which, once established, could be the first major demonstration of the advantages and the problems of inter-utility sharing.

These preliminary efforts have revealed some difficulties which need to be overcome before a link among the bibliographic utilities may be implemented.

- The four organizations that are candidates for linking in the U.S.--OCLC, RLG, WLN, and LC--have four different perspectives, making agreement difficult on issues related to linking. Each institution has its own goals and priorities, its own facilities, operating systems, and unique service capabilities. Each has its own aspirations for its future role in the library community.

- The institutions affiliated with each of the utilities have overlapping interests. Some libraries have already switched their affiliation from one utility to another, and it is unclear what, if any, effect linking might have on this "migration".

- Each bibliographic utility is in a different stage of technical development, and each is committed to the development of systems for new services it has already planned.

- The NCLIS/NBS communications protocol has not been fully tested, and there are other standards that must be developed before an online link can be implemented.

- As often happens when groups of people work hard to accomplish a goal, organizational pride sometimes makes compromise and cooperation difficult.

In addition, there has been a lack of data to support decisions about why and how the institutions should be linked. The Council on Library Resources (CLR) seeks through this study to generate data upon which decisions about linking may be made. This effort is one of several activities within CLR's Bibliographic Service Development Program (BSDP). Begun in 1978, the BSDP works with many organizations toward three goals: (a) To provide effective bibliographic services for all who need them, (b) To improve the nature and quality of bibliographic products, and (c) To stabilize the costs of many bibliographic processes in individual libraries (1). Pursuant to these goals, one of the first BSDP activities has been to explore the benefits, costs, and methods of linking.

Initial discussions between CLR staff and the BSDP Program Committee concluded that a study should be conducted to gather data on how linking would affect the utilities and their users and to provide guidance concerning alternative modes of linking that might be established. This report is one of the products of that study. An interactive computer model, called BIBLINK, was also developed. This tool was used to assess the economic impact on the utilities of linking in various ways, and it may be used by others in further evaluation of link alternatives that are judged to be promising.

---

(1) C. Lee Jones and Nancy Gwinn, "Bibliographic Service Development: A New CLR Program", Journal of Library Automation, Vol. 12, No. 2 (June 1979), pp. 116-124.

Organization of the Report

The findings and analyses of data are presented in two parts. First, three chapters examine the effects of linking on libraries and their patrons. This segment of the report addresses the general question: Is there any motivation to link the data bases of the utilities? Chapters 2, 3, and 4 provide detailed analyses of the benefits of linking for shared cataloging, interlibrary loan, and reference searching, respectively.

The second part of the report examines the technical requirements for establishing a link, analyzing a number of link alternatives. Chapter 5 discusses several of the linking methods, focusing on three promising alternatives for further analysis: (1) tape exchange of requested records, (2) online access using the "native mode" of the linked utility, and (3) online access with automatic translation of searches and responses. Chapter 6 describes the requirements for developing each of these links, and compares the alternatives on a number of quantitative and qualitative criteria.

A summary of conclusions and recommendations is presented in Chapter 7. The Appendices present supporting data developed during the study and elaborations of some topics.

CHAPTER 2

THE BENEFITS OF LINKING FOR SHARED CATALOGING

Introduction

Online shared cataloging is one of the most important services provided by the bibliographic utilities. The service enables libraries to copy bibliographic records originally produced by another library and to either adopt the record in the exact form retrieved or to modify parts of it according to local cataloging practices and preferences. This "copy cataloging" often eliminates part of the expensive and duplicative effort of original cataloging.

A survey of utility users in the U.S. has indicated the relative importance of shared cataloging activities, in terms of the proportion of time libraries spend using the online data bases for cataloging. Joseph Matthews (1) found in 1978 that OCLC and RLIN members spent approximately 75 percent of their usage hours on current cataloging and retrospective conversion. This large share of terminal time is devoted to shared cataloging because of the advantages of using an online union catalog. A study by Barbara Markuson (2) found that some of these advantages are more efficient use of staff time, faster processing, and higher quality bibliographic records for the local catalog. Thus, using the bibliographic utilities for shared cataloging offers economic benefits and can also improve service to patrons by enabling libraries to improve their local catalogs and place books on the shelves sooner.

---

(1) Joseph R. Matthews, "The Four Online Bibliographic Utilities: A Comparison", Library Technology Reports, 15:6 (November-December 1979), p. 741.

(2) Barbara Markuson, "Ohio College Library Center System: A Study of Factors Affecting the Adaptation of Libraries to Online Networks", Library Technology Reports, 12:1 (January 1976).

Linking the utilities might increase these benefits of using online union catalogs. For example, more machine-readable bibliographic records may become available to catalogers and, as a result, fewer titles would require original cataloging. The cost of cataloging might be reduced in this way. In addition, cataloging might proceed more rapidly for some titles because of the additional sources of bibliographic records. Some books may become available to the patrons sooner than if the library had waited for a record to appear on the host data base. Appendix 2A contains a more detailed discussion of the impact of linking on the timing of current title cataloging.

The rationale for linking is based on the expectation that linking would indeed increase the hit rate for cataloging at some or all of the utilities. There is no empirical evidence to confirm this, and yet solid data are needed if the increase in hit rates and the reduction in cataloging cost are to be predicted with confidence.

One of the the fundamental tasks of this study was to measure the benefits that libraries would gain in the area of cataloging if OCLC, RLIN, WLN, and LC were linked. The general approach taken was to measure hit rates on each utility for a sample of current titles and to calculate the increase in hit rate for each utility, assuming that links existed. A simple model was also used to estimate the economic benefit that would result from the increased hit rates.

Methodology

A large number of monograph titles was selected from those that were being cataloged by utility member libraries, and each title was searched once a week on each of the four data bases until an LC-MARC record appeared on each utility or until 8 weeks had passed. The resulting data were used to calculate the increase in hit rate for this sample of titles in each of the 8 weeks.

The test search was designed to shed light on some questions about the current situation for utility users and also the situation if linking had been available:

(1) What is the hit rate for current monographs on OCLC, RLIN, WLN, and LC?

(2) Do the hit rates appear to be related to data base size?

(3) What is the overlap in coverage of the four data bases?

(4) How quickly do records for current monographs become available on each data base?

(5) What would be the increase in hit rate if each utility were linked to one or more of the others?

(6) Do the incremental hit rates differ for different types of libraries? For different types of material?

Each of 20 libraries was asked to contribute a list of 50 titles that it had recently ordered or acquired. The sample of titles was restricted to current monographs because it was felt that the hit rates for these titles could be generalized to all users of a utility, while titles from retrospective conversion projects would only represent the small number of libraries involved in the study. It is likely, of course, that linking would increase hit rates for retrospective conversions, but this aspect of shared cataloging was not examined.

The participants in the study included ten libraries from OCLC and five each from RLIN and WLN. Among the twenty, there were five members of the Association of Research Libraries, five other academic libraries, five large public and five small public libraries. In addition, 250 titles were selected at random from the most recent issue of "Publishers Weekly" and added to the list of titles acquired. This was done so that the final sample for the experiment would include new books that may not have been selected yet by acquisitions librarians but which were very likely to be ordered and perhaps acquired by the time the test search began.

Each title was assigned to one of five categories: fiction, humanities, social science, science and technology, and non-English. After eliminating errors and duplicates and balancing the number of titles in each category, a final sample of 849 titles was selected. Appendix 2B shows the composition of the sample with respect to utility affiliation, type of library, and type of material.

All 849 titles were given to four people, each of whom was trained in the use of one of the four systems. OCLC was searched through an OCLC terminal in Battelle's library. RLIN and WLN were searched from dial-up terminals at Battelle. The LC-MUMS system, which provides access to new LC-MARC records, preliminary LC-MARC records from the Editorial Office, and the Automated Process Information File (APIF), was searched by a Battelle staff member using a terminal at the Library of Congress.

Searching began in December 1979 and ended in January 1980. Titles were searched on each system until an LC-MARC record was found or until they had been searched once each week for 8 weeks. After 8 weeks, the following data had been recorded for each of the four utilities:

(1) The number of full LC-MARC record hits each week

(2) The number of Cataloging in Publication (CIP) record hits each week

(3) The number of member-generated record hits each week

(4) For each title, a history of when it was first found and how its cataloging level changed over the period of the searches.

These data were used to compute for each utility the total number of hits each week out of the 849 titles. The "no-hits" on each data base were compared to determine how many titles could have been found on another utility in a given week and the level of cataloging that would have been encountered.

The limitations of this test search must be recognized. The sample of titles was not intended to be representative of all current title cataloging for the period. For this reason, the categories did not contain a representative distribution of titles. For example, the 258 titles contributed by OCLC members do not constitute a proportion of all OCLC titles that is large enough to support statistically valid inferences about the hit rates experienced by OCLC libraries in general. Similarly, the 212 titles contributed by research libraries are not a representative sample of all current titles for all research libraries using a utility. The data collected in this study do, however, indicate the relative standing of the four data bases in January 1980 with respect to current titles. The results of the exercise can thus provide a focus for further discussion and analysis of the increase in hit rates and the other advantages to be gained by linking the utilities.

Results of the Current Title Test Search

The findings of the test search and their implications are detailed in the following sections. These results are presented in three parts:

(1) Current title hit rates on each utility
(2) Increases in hit rates when utilities are linked
(3) Estimates of the economic benefits resulting from the increases in hit rates.

Current Title Hit Rates

The test search was conducted over a period of 8 weeks in order to reveal the change in hit rate across time for the four data bases. Many libraries routinely hold items longer than 2 months while waiting for acceptable copy to appear in a data base, but the 8-week duration of the test search does reveal the availability of records during the period when most libraries conduct the first few searches for a title. Figure 2-1 shows the hit rate on each data base in the search experiment, and Figure 2-2 is a graph of these data. The numbers do not reflect the appearance of an upgraded record after the first hit on the title was recorded.

Data Base Overlap. Figure 2-3 shows that there is considerable overlap in coverage among the bibliographic data bases. Approximately 80 percent of the titles were found on all four of the data bases by the end of

FIGURE 2-1. HIT RATES FOR CURRENT TITLES

Results of a test search of 849 current titles in December 1979 and January 1980.

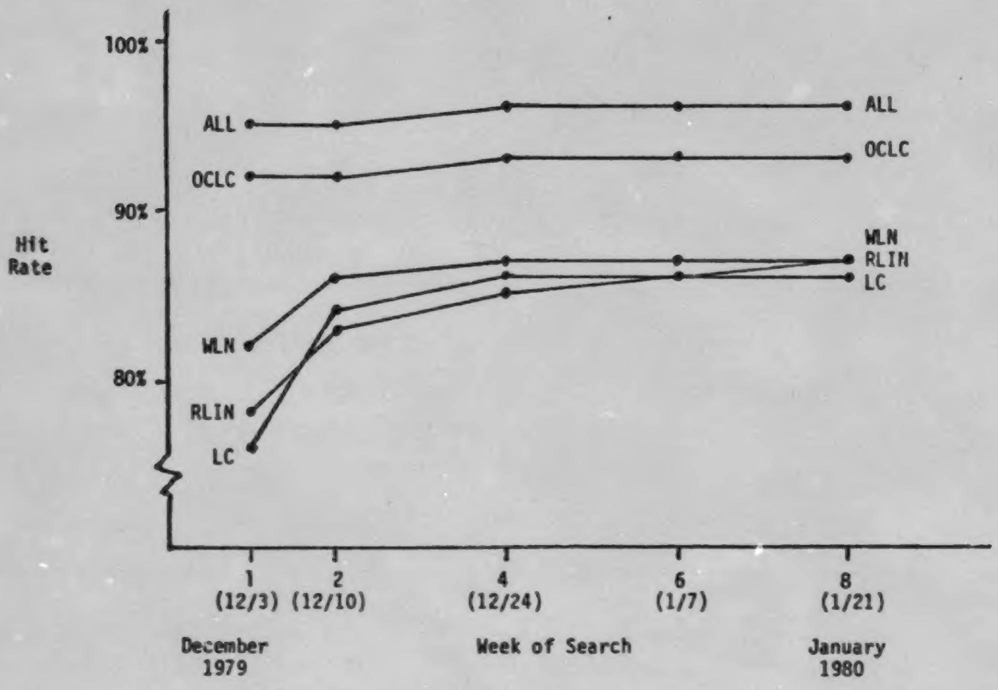
Numbers represent the percentage of titles for which a record was found on a given data base.

Data Base	Week 1	Week 2	Week 4	Week 6	Week 8
OCLC	92%	92%	93%	93%	93%
RLIN	78%	83%	85%	86%	87%
WLN	82%	86%	87%	87%	87%
LC	76%	84%	86%	86%	86%
All	95%	95%	96%	96%	96%

FIGURE 2-2. GRAPH OF HIT RATES FOR CURRENT TITLES

Results of a test search of 849 current titles in December 1979 and January 1980.

Points plotted are hit rates from Figure 2-1.



the eighth week, and all but a few of these were found on each data base after 2 weeks of searching. The differences in the hit rates shown for each utility in Figure 2-1, then, are due to the different availability on each data base of records for the roughly 20 percent of the titles that were not held in common by all.

FIGURE 2-3. OVERLAP OF DATA BASE COVERAGE FOR CURRENT TITLES

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles for which a record was found on the stated number of data bases.

Number of Items Found In:	Week 2	Week 4	Week 6	Week 8
All four data bases	76.6%	79.2%	80.2%	80.3%
Three data bases	8.1%	5.8%	5.1%	5.6%
Two data bases	4.8%	4.7%	5.1%	4.9%
One data base	5.8%	5.8%	5.2%	4.9%
None of the data bases	4.7%	4.5%	4.4%	4.3%
TOTAL	100%	100%	100%	100%

LC-MARC Hit Rates. Figure 2-4 breaks the Week 8 hit rate for each utility into its LC-MARC and non-LC-MARC components. The routine validation of records from the MARC tapes took more time at RLIN than at OCLC at the time of the test search, and this may account for the difference in their LC-MARC hit rates. WLN's LC-MARC figure is abnormally low. WLN loaded the

MARC tapes only once during the test search because they were performing data base maintenance activities during that period. Under normal circumstances, the LC-MARC hit rate on WLN would probably have been 10 to 15 percentage points higher. The high LC-MARC hit rate on the Library of Congress system reflects the presence of undistributed MARC records on that data base. These anomalous conditions make it risky to draw conclusions from these data. It should be recognized, though, that if each data base had had the same LC-MARC availability during the period, the total hit rate figures might have been somewhat different. For example, if RLIN and WLN had had all of the post-1971 LC-MARC records that OCLC had at the time, their current title hit rates could have been up to 4 percentage points higher.

FIGURE 2-4. HIT RATES ON LC-MARC RECORDS

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles for which a record was found on a given data base.

Data Base	Week 8 LC-MARC Hit Rate	Week 8 Non-LC-MARC Hit Rate	Week 8 Total Hit Rate
OCLC	65%	28%	93%
RLIN	61%	26%	87%
WLN	50%	37%	87%
LC	72%	14%	85%

Hit Rates by Title Category. Figures 2-5, 2-6, and 2-7 show the eighth-week hit rates for subgroups of the sample. OCLC had the highest hit rate in every category; more titles were found on OCLC than on any other data base, regardless of the source of the title or the type of material. A particularly notable point is that the hit rate for titles contributed by research libraries was higher by 10 percentage points on OCLC than on RLIN. The difference in the OCLC and RLIN hit rates for non-English titles is even greater, suggesting that at the time of the test search, OCLC had a decided edge over RLIN in its usefulness to research libraries. Two factors must be borne in mind when interpreting these findings, though. First, OCLC retains the records that were created by RLG libraries previously affiliated with OCLC, and much of this material had not been added to RLIN's data base at the time of the study. Second, several research libraries have become RLG members since this study was conducted, and several more will likely join RLG within the next several months. The growth of RLG and continued retrospective conversion by all RLIN users is likely to bring RLIN closer to OCLC's hit rate for material cataloged by research libraries.

Figure 2-5 shows that the titles contributed by public libraries had a higher hit rate on each of the data bases than those contributed by either academic or research libraries. This result is not surprising, since the current title cataloging at public libraries normally contains more material for which an LC-MARC record is available. It should also be noted that after 8 weeks the titles contributed by RLIN and WLN members had higher hit rates on OCLC than on their own data bases, as shown in Figure 2-6. This is an indication of the benefit that most libraries would have gained from linking with OCLC at the time of the test search. The hit rates for each type of material except non-English titles were reasonably close for

FIGURE 2-5. HIT RATES FOR CURRENT TITLES--TITLES CATEGORIZED BY TYPE OF CONTRIBUTING LIBRARY

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles for which a record was found on a given data base.

Data Base	Public Titles (n=236)	Academic Titles (n=159)	Research Titles (n=212)	All Titles (n=849*)
OCLC	97%	98%	82%	93%
RLIN	88%	87%	72%	87%
WLN	92%	84%	70%	87%
LC	89%	84%	69%	86%

* Includes 242 titles selected from the Weekly Record.

FIGURE 2-6. HIT RATES FOR CURRENT TITLES--TITLES CATEGORIZED BY UTILITY OF CONTRIBUTING LIBRARY

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles for which a record was found on a given data base.

Data Base	OCLC Titles (n=258)	RLIN Titles (n=176)	WLN Titles (n=173)	Weekly Records (n=242)	All Titles (n=849)
OCLC	91%	85%	85%	99%	93%
RLIN	86%	76%	76%	98%	87%
WLN	79%	74%	74%	99%	87%
LC	81%	74%	74%	99%	86%

FIGURE 2-7. HIT RATES FOR CURRENT TITLES--TITLES CATEGORIZED BY TYPE OF MATERIAL

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles for which a record was found on a given data base.

Data Base	Fiction Titles (n=87)	Humanities Titles (n=219)	Soc Sci Titles (n=219)	Sci/Tech Titles (n=224)	Non-English Titles (n=100)	All Titles (n=849)
OCLC	94%	97%	97%	91%	71%	93%
RLIN	90%	92%	93%	88%	55%	87%
WLN	92%	94%	94%	90%	47%	87%
LC	87%	92%	89%	88%	60%	86%

all four data bases. Figure 2-7 shows that OCLC's hit rate for non-English titles was 71 percent, compared to 55 percent for RLIN and 47 percent for WLN.

Incremental Hit Rates

Figure 2-8 lists the linked hit rates for each combination of data bases at 2-week intervals. Figure 2-9 shows Week 8 linked hit rates for all breakdowns of the sample. The incremental hit rate from linking may be calculated from either table by selecting a combination of data bases and subtracting the number shown under the heading "One Data Base" from that shown for the desired combination. For example, if WLN were to be linked to all three other data bases, then the hit rate for WLN users would be 96 percent, an increase of 9 percent over the unlinked hit rate of 87 percent for WLN libraries.

Figure 2-10 shows the incremental hit rates for each utility in 2-week intervals, and Figure 2-11 summarizes these increments for each utility. At the end of 8 weeks, LC users had the largest gain--10 percentage points added to their current title hit rate. RLIN and WLN libraries would have gained 9 points if linked to the other utilities, and OCLC members could have added 3 percentage points to their hit rate.

These hit rate increases represent the maximum advantage that could be gained by a library searching all linked data bases in the given week. This assumes that, with an online link, a library would always search for and would always find the records held elsewhere that are not found on

FIGURE 2-8. HIT RATES FOR ALL CURRENT TITLES ON COMBINATIONS OF DATA BASES

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles for which a record was found on the combination of data bases shown.

	Week 2	Week 4	Week 6	Week 8
One Data Base				
OCLC	92	93	93	93
RLIN	83	85	86	87
WLN	86	87	87	87
LC	84	86	86	86
Two Data Bases Linked				
OCLC-RLIN	94	94	94	94
OCLC-WLN	94	94	94	94
OCLC-LC	94	94	94	94
RLIN-WLN	90	90	91	91
RLIN-LC	89	90	90	90
WLN-LC	90	90	90	91
Three Data Bases Linked				
OCLC-RLIN-WLN	95	95	95	96
OCLC-RLIN-LC	94	95	95	95
OCLC-WLN-LC	95	95	95	95
RLIN-WLN-LC	92	92	92	93
Four Data Bases Linked				
OCLC-RLIN-WLN-LC	95	96	96	96

FIGURE 2-9. WEEK 8 HIT RATES FOR CURRENT TITLES ON COMBINATIONS OF DATA BASES

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage of titles in each category for which a record was found on the combination of data bases shown.

	All Titles (n=849*)	OCLC Titles (n=258)	RLIN Titles (n=176)	WLN Titles (n=173)	Public Titles (n=236)	Academic Titles (n=159)	Research Titles (n=212)	Fiction Titles (n=87*)	Humanities Titles (n=219*)	Soc Sci Titles (n=219*)	Sci/Tech Titles (n=224*)	Non-English Titles (n=100*)
One Data Base												
OCLC	93	91	85	85	97	93	82	94	97	97	91	71
RLIN	87	86	76	76	88	87	72	90	92	93	88	55
WLN	87	79	74	74	92	84	70	92	94	94	90	47
LC	86	81	74	74	89	84	69	87	92	89	88	60
Two Data Bases Linked												
OCLC-RLIN	94	93	87	96	97	96	85	94	98	99	92	74
OCLC-WLN	94	92	87	97	98	96	85	94	98	98	93	72
OCLC-LC	94	93	86	95	97	94	84	94	97	97	92	76
RLIN-WLN	91	91	81	97	94	90	80	92	95	95	94	62
RLIN-LC	90	91	82	91	92	89	78	91	96	95	90	67
WLN-LC	91	82	80	97	92	88	77	92	95	95	92	63
Three Data Bases Linked												
OCLC-RLIN-WLN	96	94	88	98	98	96	87	94	99	99	97	75
OCLC-RLIN-LC	95	95	88	96	97	96	86	94	98	99	97	78
OCLC-WLN-LC	95	94	87	98	98	95	87	94	98	98	97	77
RLIN-WLN-LC	93	89	85	98	95	91	84	92	98	96	95	72
Four Data Bases Linked												
OCLC-RLIN-WLN-LC	96	95	93	98	98	96	88	94	99	99	97	79

* Includes titles selected from the Weekly Record.

FIGURE 2-10. INCREMENTAL HIT RATES FOR ALL CURRENT TITLES

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage points added to the hit rate when utilities are linked.

	Host Utility	Incremental Hit Rate When Linked To:				
		OCLC	RLIN	WLN	LC	All
Week 2	OCLC	--	2%	2%	2%	3%
	RLIN	11%	--	7%	6%	12%
	WLN	8%	4%	--	4%	9%
	LC	10%	5%	6%	--	11%
Week 4	OCLC	--	1%	1%	1%	3%
	RLIN	9%	--	5%	5%	11%
	WLN	7%	3%	--	3%	9%
	LC	8%	4%	4%	--	10%
Week 6	OCLC	--	1%	1%	1%	3%
	RLIN	8%	--	5%	4%	10%
	WLN	7%	4%	--	3%	9%
	LC	8%	4%	4%	--	10%
Week 8	OCLC	--	1%	1%	1%	3%
	RLIN	7%	--	4%	3%	9%
	WLN	7%	4%	--	4%	9%
	LC	8%	4%	5%	--	10%

FIGURE 2-11. INCREMENTAL HIT RATES FOR ALL CURRENT TITLES WHEN ALL DATA BASES ARE LINKED

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage points added to the hit rate when utilities are linked.

Utility	Week 2	Week 4	Week 6	Week 8
OCLC	3%	3%	3%	3%
RLIN	12%	11%	10%	9%
WLN	9%	9%	9%	9%
LC	11%	10%	10%	10%

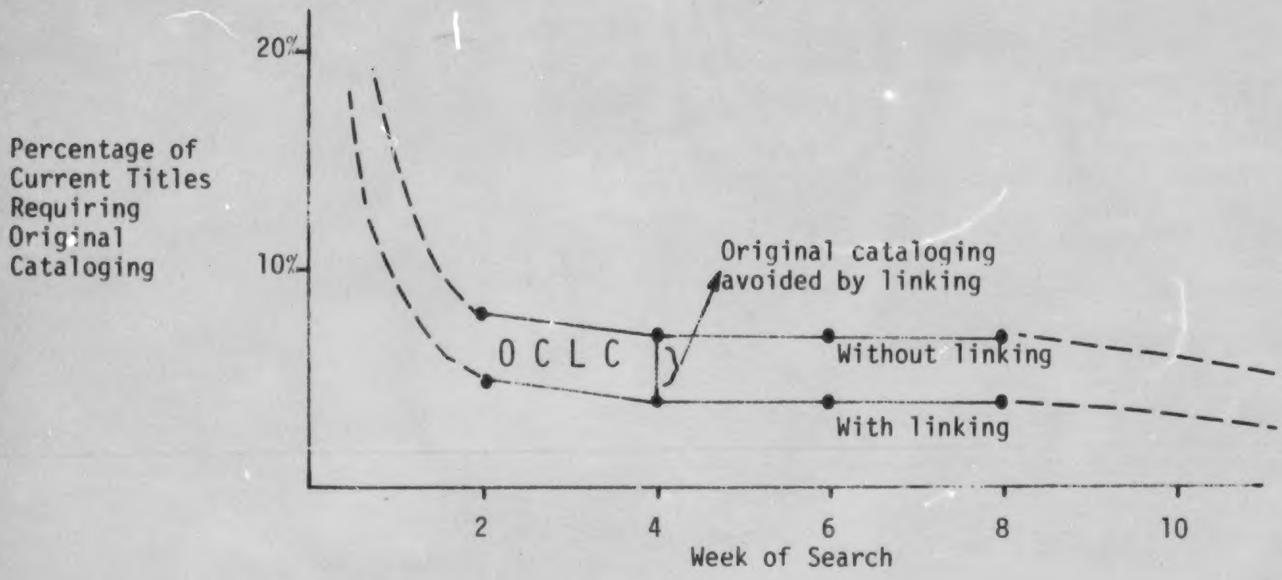
the host data base. If catalogers do not routinely search until they hit the record available elsewhere, or if fewer than three other data bases are searched, then the increase in hit rate will be somewhat lower.

OCLC's incremental hit rate from linking is a steady 3 percentage points from Week 2 through Week 8. WLN's 9 point increase is also stable throughout the period. The advantage of a four-way link for RLIN users drops from 12 to 9 percentage points over the period, indicating that records for some of the sample titles were added to the RLIN data base during the search test. LC's advantage moves from 11 to 10 percentage points.

Figure 2-10 shows that the largest part of the hit rate increase for WLN, RLIN, and LC comes from linking to OCLC. A connection between RLIN and WLN alone would benefit the users of each by about 4 percentage points. OCLC libraries would not experience an increase in hit rate of more than 1 percentage point when linked with any other data base, and linking with LC would add no more than 4 points to any library's hit rate by the end of 2 months. These data indicate that RLIN, WLN, and LC users can benefit most by linking to OCLC's data base. OCLC libraries would have a smaller increase in hit rate, though this might lead to a significant reduction in original cataloging for some of them.

Figure 2-12 illustrates these results in graphical form. Each graph shows the reduction in original cataloging that users of each utility could have realized from linking in early 1980. The reduction in original cataloging is assumed to be equivalent to the "no-hit rate" for a given week. In other words, the decrease in the percentage of books requiring

FIGURE 2-12. ORIGINAL CATALOGING AVOIDED BY LINKING



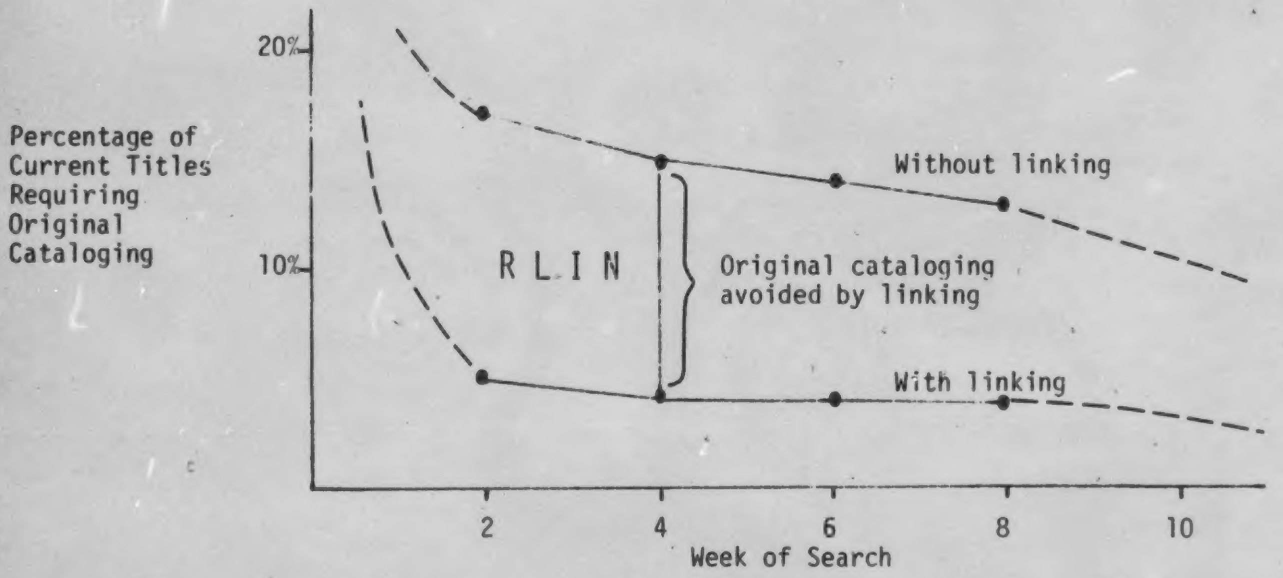
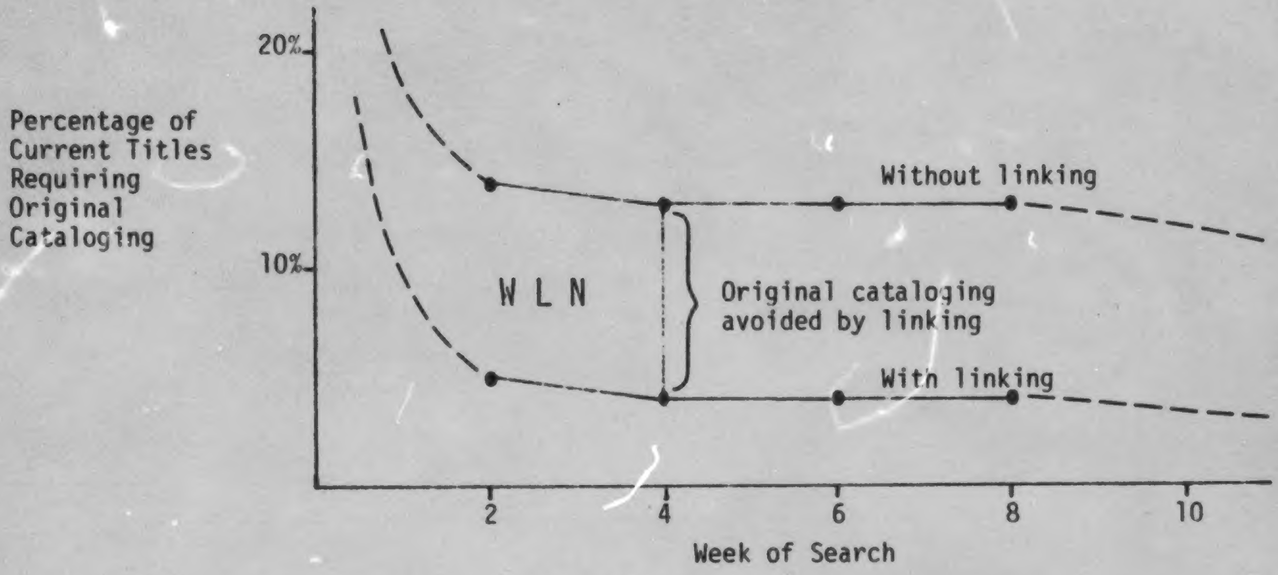
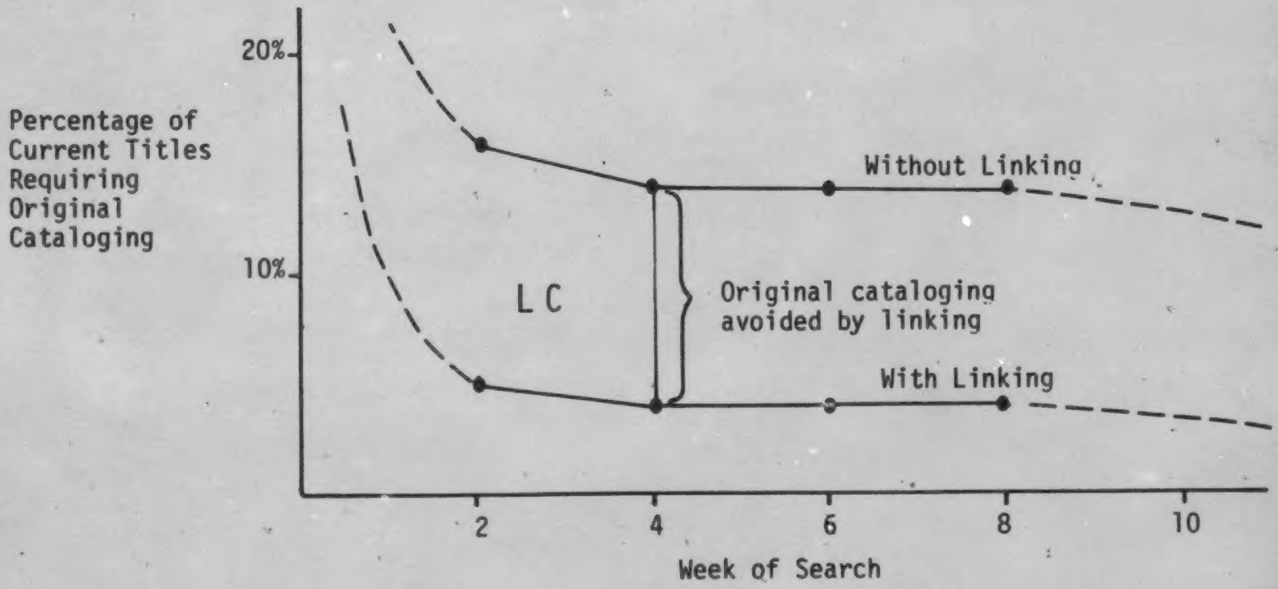


FIGURE 2-12
(Continued)





original cataloging is considered equivalent to the increase in the hit rate shown in Figure 2-11.

Incremental Hit Rates by Title Category. Figures 2-13 through 2-15 present the eighth-week incremental hit rates for titles contributed by members of each utility, for titles contributed by each type of library, and for non-English titles. In Figure 2-13, the advantage to RLIN and WLN users of linking to OCLC is again apparent. Eleven percent of the titles contributed by RLIN libraries were found on OCLC but not on RLIN. For WLN, this figure was 23 percent. Figure 2-14 suggests that, in January 1980, linking could have been of greatest value to research libraries, enabling many of them to increase their current title hit rates by 15 or 20 percentage points. The increments for public and academic libraries were smaller but still significant. Figure 2-15 shows that linking would have improved the hit rates on non-English titles by 8 percentage points for OCLC members and by 20 to 30 points for users of other utilities.

The Economic Benefits of Linking for Shared Cataloging

The main benefit of linking for shared cataloging is the opportunity to catalog a title by using a bibliographic record produced elsewhere, avoiding part of the effort, and thus part of the cost, of original cataloging. Figure 2-16 illustrates an approach to estimating the cost avoided. This table is a general model for calculating the annual cost of original cataloging avoided for current titles by a single library linked to any combination of the bibliographic data bases discussed in this report. The Week 4 hit rates are used in this analysis because, as shown in Figure 2-2, most of the titles used in the search test had hit records by that

FIGURE 2-13. INCREMENTAL HIT RATES FOR CURRENT TITLES CONTRIBUTED BY LIBRARIES FROM EACH UTILITY

Results of a test search of 849 current
titles in December 1979 and January 1980.

Numbers represent the percentage points added
to the hit rate when utilities are linked.

Utility of Libraries Contributing Titles	Incremental Hit Rate When Linked To:				
	OCLC	RLIN	WLN	LC	All
OCLC (n=258)	--	2%	1%	2%	4%
RLIN (n=176)	11%	--	5%	6%	17%
WLN (n=173)	23%	23%	--	23%	24%

FIGURE 2-14. INCREMENTAL HIT RATES FOR CURRENT TITLES CONTRIBUTED BY THREE TYPES OF LIBRARIES

Results of a test search of 849 current titles in December 1979 and January 1980.

Numbers represent the percentage points added to the hit rate when utilities are linked.

PUBLIC LIBRARIES (n=236)

Host Utility	Incremental hit rates when linked to:				
	OCLC	RLIN	WLN	LC	All
OCLC	--	0	1%	0	1%
RLIN	9%	--	6%	4%	10%
WLN	6%	2%	--	0	6%
LC	8%	3%	3%	--	9%

ACADEMIC LIBRARIES (n=159)

Host Utility	Incremental hit rates when linked to:				
	OCLC	RLIN	WLN	LC	All
OCLC	--	3%	3%	1%	3%
RLIN	9%	--	3%	2%	9%
WLN	12%	6%	--	4%	12%
LC	10%	5%	4%	--	12%

RESEARCH LIBRARIES (n=212)

Host Utility	Incremental hit rates when linked to:				
	OCLC	RLIN	WLN	LC	All
OCLC	--	3%	3%	2%	6%
RLIN	13%	--	8%	6%	15%
WLN	15%	10%	--	7%	18%
LC	15%	9%	8%	--	19%

FIGURE 2-15. INCREMENTAL HIT RATES FOR CURRENT NON-ENGLISH TITLES

Results of a test search of 849 current
titles in December 1979 and January 1980.

Numbers represent the percentage points added
to the hit rate when utilities are linked.

n = 100

Host Utility	Incremental Hit Rate When Linked To:				
	OCLC	RLIN	WLN	LC	All
OCLC	--	3%	1%	5%	8%
RLIN	19%	--	7%	12%	24%
WLN	25%	15%	--	16%	32%
LC	16%	7%	3%	--	19%

point. In the absence of empirical data, Battelle has supplied assumptions about the magnitude of some of the factors used in the model. Individual libraries may tailor the calculations to their specific conditions by supplying different data and assumptions. A blank version of Figure 2-16 is provided as Appendix 2C.

It must be emphasized again that these estimates of cost avoidance are based on hit rates from a relatively small sample of current titles searched during December 1979 and January 1980. Changing conditions since then could have altered these hit rates. Moreover, while hit rates are used here as the basis for assessing the economic benefits, other factors need to be considered in analyzing the full benefits from linking the utilities. The study did not attempt to assess the non-quantitative effects of such factors as differences in the quality of records maintained by the utilities or differences in the quality of services, even though these are likely to be important issues with regard to linking.

FIGURE 2-16. GENERAL MODEL FOR CALCULATING ANNUAL CATALOGING COSTS AVOIDED BY LINKING

	OCLC	RLIN	WLN
1. Current titles cataloged this year, per library	5,000(1)	12,500(2)	5,700(3)
2. Reduction in percentage of titles requiring original cataloging (Week 4 incremental hit rate) (4)	.03	.11	.09
3. Potential number of titles for which original cataloging may be avoided this year	150	1,375	513
4. Fraction of these for which original cataloging is actually avoided (5)	.5	.5	.5
5. Number of titles for which original cataloging is actually avoided this year	75	688	257
6. Cost avoided per title if copy cataloging is done instead of original cataloging(5)	$15	$15	$15
7. Total cost of original cataloging avoided this year, per library	$1,125	$10,320	$3,855
8. Cost avoidance per title cataloged (Line 7 ÷ Line 1)	$.23	$.83	$.68
9. Approximate number of affiliated institutions performing online cataloging	2,100	80	59
10. Aggregate cost avoidance for the network (Line 7 x Line 9)	$2,362,500	$825,600	$227,445

(1) Assumes 80% of titles cataloged are current titles. OCLC estimates 12 million first-time uses plus 1 million new records added this year. 13 million x .80 = 10.4 million current titles cataloged. 10.4 million titles ÷ 2,100 OCLC members = roughly 5,000 current titles per member.

(2) Assumes 80% of titles cataloged are current titles. RLG estimates 1.25 million titles cataloged this year. 1.25 million x .80 = 1 million current titles cataloged. One million titles ÷ 80 institutions that use RLIN for cataloging = 12,500 current titles per library.

(3) Assumes 80% of titles cataloged are current titles. WLN estimates 420,000 titles cataloged this year. 420,000 x .80 = 336,000 current titles cataloged. 336,000 current titles ÷ 59 WLN participants = about 5,700 current titles per library.

(4) Results of a test search of 849 current titles in December 1979 and January 1980. See Figure 2-11.

(5) Battelle estimate.

A Model for Calculating Cataloging Cost Avoided

Line 1. Figure 2-16 begins with an assumption about the number of current titles cataloged per year at an "average" library affiliated with each utility.

OCLC projects a total cataloging activity this year of about 13,000,000 titles, for an average activity per member library of about 6,200 titles. It was assumed that 80 percent of this cataloging will be for current titles, or approximately 5,000 current titles this year for each OCLC library.

Similarly, RLG estimates a total cataloging activity of 1,250,000 titles in the current year, for an average volume per RLIN library of 15,600 titles. Again, if it is assumed that 80 percent of this activity is for current title cataloging, the average RLIN user will catalog 12,500 current titles this year.

Finally, WLN projects a load of 420,000 titles, for an average of 7,100 titles per member. Current title cataloging, at 80 percent of this total, is about 5,700 titles per library.

Line 2. The next step is to calculate the portion of these titles for which original cataloging could be avoided after 4 weeks if all four data bases were linked. Line 2 in Figure 2-16 contains the increase in hit rate when a utility is linked to other data bases. For example, linking to all three others raises OCLC's hit rate by 3 percentage points.

Line 3. When the incremental hit rate is multiplied by the number of titles cataloged in the year, the result is the maximum number of titles for which original cataloging could be avoided.

Line 4. Two assumptions underly the figure used on Line 4. First, it is assumed that all searches that do not find an acceptable record in the host data base will be repeated on linked data bases until a record is found or until all of the linked data bases have been searched. That is, the persistence factor is assumed to be 1. In practice, this persistence factor for linked searching is surely less than 1, but it will vary widely from library to library and probably from title to title at each library. The effect of this simplifying assumption is that the total cost avoidance calculated in Figure 2-16 represents the maximum benefit possible if all of the available resources are used.

The second assumption concerns the acceptability of records found in linked data bases. It is unlikely that all of the records found elsewhere will be acceptable to a cataloger. Most of the records found on other data bases would be member-generated records contributed by that utility's libraries, and some catalogers will choose to do their own original cataloging rather than accept the work of other libraries. In the absence of data concerning this acceptance factor, a conservative estimate was used: only half of the records found through linking would actually be used to avoid original cataloging. A different assumption may be easily substituted by others using this model for their own analyses.

Line 5. The product of Lines 3 and 4 is the number of titles for which original cataloging is avoided.

Line 6. The cost of doing original cataloging is higher than that for using a shared record. The difference in cost will vary substantially from library to library, depending on local cataloging procedures. A conservative estimate of $15 per title was used in these calculations. Discussions with research libraries reveal cost differentials of $20 to $45, while most public libraries probably experience much less savings. Research libraries may wish to insert a higher figure when using this model.

Line 7 and Line 8. Multiplying the cost avoided per title by the number of titles for which original cataloging is avoided results in the estimated total cost avoided by an average library in the year. These calculations indicate that linking for shared cataloging would be beneficial to many libraries, enabling them to avoid a significant portion of the cost of cataloging current titles. An average RLIN user might avoid more than $10,000 in cataloging costs this year if access to the other data bases were available, a savings of $.83 per current title cataloged by the library. WLN libraries could avoid an average of nearly $4,000 each, or $.68 per title cataloged. The benefit to the average OCLC library is smaller still, reflecting the higher initial hit rate on OCLC. OCLC libraries would average about $1,100 of original cataloging costs avoided this year, or $.23 per title cataloged. The cost avoided by OCLC's large libraries will be much greater than $1,100, though.

While these figures are only estimates, they show the magnitude of the economic benefit for many libraries when the utilities are linked. The cost avoided per title cataloged may prove to be a more useful number than the total cost avoided by an average library. Using the unit savings shown on Line 8 in Figure 2-16, any library can very quickly estimate the annual cataloging costs it could avoid through linking. This is done by multiplying the cost avoided per title by the number of current titles the library catalogs in a year.

Figure 2-16 also provides an estimate of the aggregate cost avoidance for the user community of each utility. WLN libraries could avoid about $227,000 in cataloging costs this year if they had access to the three other data bases. RLIN's aggregate savings is about $825,000, including the cost that could be avoided by RLIN users that are not members of RLG. OCLC's aggregate cost avoidance would be around $2.36 million this year.

The cost avoidance figures resulting from these calculations are very sensitive to changes in the values assigned to some of the variables. For example, if the incremental hit rate on Line 2 were 1 percentage point lower, and if the acceptance factor on Line 4 were .4 instead of .5, then the original cataloging cost avoided by an OCLC library would be $600 a year, just over half of that projected in Figure 2-16. On the other hand, a less conservative value for the cost per title on Line 6 would increase the final figure. For instance, each additional dollar of cost adds $688 to the cost avoided by RLIN users.

The Cost of Linking for Shared Cataloging

The cataloging cost avoided by linking will be partially offset by charges that utilities would impose on non-member users. This study did not attempt to predict the rate structure that would exist if a four-way link were established. However, several types of charges might be assessed by utilities. There is likely to be a fee for each search performed on a linked data base. In addition, there may be a "take home" fee, a charge levied when a library copies a record from a linked data base and enters the copy into its host data base. (Note that if a record is claimed from another data base, then there may be no more linked search fees associated with the title. The record is always found on the host data base, and only those who do not find it acceptable would want to search elsewhere.)

The utilities might also institute a royalty charge, similar to OCLC's first-time use fee. Under this arrangement, a library that uses a record that was obtained from another data base would have to pay the other utility a fee for copying from it. Finally, there might be a fixed monthly or annual charge for the opportunity to do linked searching, perhaps combined with one of the record-based charges already mentioned.

The average cost of a linked search cannot be predicted at this time, but an estimate can be made of the unit cost per search that would entirely offset the cost avoided by linking. Figure 2-17 shows the number of linked searches that a library would perform in a year, given the assumptions made in Figure 2-16. In OCLC's case, 350 titles are searched via a link. The cost of obtaining the data, no matter how the fees are assessed, cannot exceed about $3.21 per title, or linking will cost more than it saves. The critical cost per title for WLN is about $5.20, while

FIGURE 2-17. CALCULATION OF THE COST PER LINKED SEARCH THAT WOULD OFFSET THE COST AVOIDED

	OCLC	RLIN	WLN
1. Current titles searched in the year	5,000(1)	12,500(2)	5,700(3)
2. No-hit rate on home data base (Week 4)(4)	.07	.15	.13
3. Number of titles to be searched on other data bases in the year	350	1,875	741
4. Total original cataloging cost avoided in the year by linking (from Figure 2-16)	$1,125	$10,320	$3,855
5. Approximate cost per linked search that would offset the cost avoidance from linking (Line 4 ÷ Line 3)	$3.21	$5.50	$5.20

(1) Assumes 80% of titles cataloged are current titles. OCLC estimates 12 million first-time uses plus 1 million new records added this year. 13 million x .80 = 10.4 million current titles cataloged. 10.4 million titles ÷ 2,100 OCLC members = roughly 5,000 current titles per member.

(2) Assumes 80% of titles cataloged are current titles. RLG estimates 1.25 million titles cataloged this year. 1.25 million x .80 = 1 million current titles cataloged. One million titles ÷ 80 institutions that use RLIN for cataloging = 12,500 current titles per library.

(3) Assumes 80% of titles cataloged are current titles. WLN estimates 420,000 titles cataloged this year. 420,000 x .80 = 336,000 current titles cataloged. 336,000 current titles ÷ 59 WLN participants = about 5,700 current titles per library.

(4) Results of a test search of 849 current titles in December 1979 and January 1980. See Figure 2-1.

RLIN libraries could pay about $5.50 per linked search before the cost avoided is entirely offset. This is not to suggest, of course, that libraries would be willing to pay these prices for linked searching. If linking is to be considered worthwhile on economic grounds, then the cost of linking cannot be so high that much of the cost avoided is replaced by the cost of using the resources of other data bases.

Other Possible Benefits from Linking

There are two other ways in which catalogers might benefit from linking the utilities. In both cases, the benefit would take the same form as that received from an increased hit rate--avoidance of some of the cost of cataloging a given number of titles.

First, linking might make available a more desirable form of cataloging than that available from the host utility. An LC-MARC record might become available earlier elsewhere, and a cataloger could save effort by copying that record instead of copying the original record found on the host data base. For current titles, however, this benefit is apt to be quite small. If the utilities update their data bases promptly with the MARC tapes received each week from the Library of Congress, then there should not be a significant difference between them in the availability of LC-MARC records. (If a link were available today, RLIN users could have access to the pre-1972 LC-MARC records that are not now in the RLIN data base, but RLG plans to add these records in the near future.)

The more important benefit in this area is that linking could make available an original record that is more desirable for copy cataloging than the member-generated record found on the host data base. For example, an academic library using one utility might in some cases wish to use a record from another data base that was created by a research library instead of a record entered in the host data base by a public library. A cost avoidance would result in those cases where much modification would have been necessary in copying the host record, while the record obtained from another data base required less work.

Linking for Authority Data

The second area of potential benefit is shared access to authority records. The catalog search experiment was designed to gauge the benefits to catalogers of increased access to bibliographic files only. The same sort of advantage might be gained if authority files were linked. The Linked Authority System Project involving RLG, WLN, and LC is a major effort to reap this benefit by establishing an online authority link.

The cost of establishing authorized headings is substantial for many large libraries. Some research libraries report that authority work makes up as much as 35 percent of the cost of cataloging. When a suitable authority record cannot be found on the host data base, it may be worthwhile to search the files of other utilities as well. If a record is available on the linked data base, then the cataloger does not have to do all of the original authority work. Just as linking can enable catalogers to avoid original cataloging for some titles, the number of original authority records created might also be reduced, with an accompanying reduction in costs that could be very significant.

Appendix 2D discusses some of the issues surrounding the linking of authority files and the establishment of a nationwide authority network.

CHAPTER 3

THE BENEFITS OF LINKING FOR INTERLIBRARY LOAN

Introduction

This chapter presents an analysis of the benefits that libraries might gain in the area of interlibrary loan if the bibliographic utilities were linked. The emphasis has been placed on the use of a utility for verifying borrowing requests (that is, requests from a library's own patrons) and locating the items and the prospective lenders. The study did not evaluate the technical requirements for linking RLIN, WLN, and LC users to the OCLC ILL subsystem for electronic transmission of requests.

Interviews with ILL librarians indicated that the bibliographic utilities are not often used in processing the requests received from other libraries. In most cases, it is assumed that the request has already been verified by the requesting library, and the task is not repeated at the receiving library. Linking might affect lending, however, by increasing the number of requests a library receives and thereby increasing the cost of ILL processing. This issue was not examined in detail in this study, and more research is required before linking's impact on high-performance lenders may be adequately assessed.

ILL librarians interviewed for this study reported that using a utility has had an important effect on the processing of borrowing requests. The online union catalog has become the first-used source of verification for many libraries, particularly for material published after 1968, when the first LC-MARC records were produced. The utility data base is also the first source used in locating the item that has been requested and in identifying prospective lenders. In both cases, the additional information available in the online union catalog enhances ILL performance by improving the ability to verify and locate.

In addition to providing easy access to a great many records and holdings statements, use of an online union catalog also saves time on verification and location. The librarians interviewed reported a significant reduction in the average amount of time needed to process a borrowing request since they began using a utility. This reduction in staff time means that some of the unit cost of processing a request may be avoided when a utility is used, an important benefit of utility affiliation. Appendix 3A discusses in greater detail the impact of utility use on ILL costs.

Linking the utilities might improve the individual library's ILL performance still further by providing access to even more records and holdings statements. Several benefits of linking are possible in the areas of verification and location:

Verification Benefits

- The "verification hit rate" might be higher. If more bibliographic records are available to a library, there may be a greater chance that a record will be found for a given title.

- The greater number of records available to use for verification may mean that fewer requests remain unverified. This can improve performance, since more requests could be sent to the potential lenders.

- Some staff time may be saved if a second data base is used for verification when a request cannot be verified on the host data base. This can mean that some of the cost of processing a borrowing request is avoided.

- For OCLC users, verification might be improved because of the search command structure of other data bases. For example, errors in the first few letters of title keywords can cause a record not to be found on OCLC. The correct citation might be found on RLIN or WLN, though, using the keyword searching feature, and the title could then be searched again on OCLC to locate the item.

Location Benefits

- The availability of a greater number of records increases the chances that at least one non-LC holding will be listed for a particular title. The frequency of these hits may be referred to as the "holdings hit rate" to distinguish it from the verification hit rate.

- The availability of more records and holding statements means that fewer requests are sent out "blind"-- that is, without knowing whether the library actually holds the requested item. Requests sent to known owners may have a greater probability of being filled, and so the additional holdings information from the linked data base might improve the fill rate.

- The increased number of holding symbols may allow the library to select a potential lender who is more likely to loan the item than those listed on the host data base. This may increase the first time fill rate, fewer requests being returned unfilled. Some of the cost of reissuing unfilled requests may be avoided.

- When no suitable holdings statements are found on the host data base, staff time may be saved by searching a second data base instead of other location sources. As is the case for time saved in verification, this can translate into an avoidance of some of the cost of processing a request.

- The availability of more holdings information may redistribute the ILL load throughout the library community, easing the burden on some large libraries that now serve as the primary source for many other libraries. If holdings information became more widely available through linking, some libraries might receive requests they would not have received otherwise, and some net lenders may experience a drop in their lending volume because of this shift. (Of course, the opposite is also possible. Libraries with high lending volumes could receive many more requests from libraries that learned of their holdings through linking.)

There are no published studies to demonstrate that these supposed advantages of linking would actually occur, and fresh data had to be collected so that the magnitude of some of these benefits could be assessed. A large number of monograph borrowing requests was searched on each data base to see how the utilities compared, at a point in time, in the availability of records and holdings information for these titles. Not all of the possible benefits outlined above could be examined in this manner, and the study was focused on a few important questions:

(1) What is the "verification hit rate" for libraries using OCLC, RLIN, and WLN? That is, how many of the titles searched were found?

(2) What is the "holdings hit rate" for each data base? That is, how many of the titles searched found a record with at least one non-LC holding statement?

(3) The increase in hit rate is one measure of the potential benefit of linking. How would the hit rates increase if users of one data base were given access to the resources of other data bases?

(4) Do the increases in hit rate differ for different types of libraries, different types of material, or for books of differing publication dates?

(5) Are ILL requests generally for material that is held by only a few of the libraries that use the same utility as the requesting library? How many of the requested items are "widely held" among libraries using another utility? The results of this inquiry are presented in Appendix 3B.

Methodology

A sample of 277 libraries was selected, and each library was asked to contribute outgoing ILL requests from a 2-week period (November 19-December 3, 1979). These libraries were all users of one of the three utilities and included research, large and small academic, and large and small public libraries. Responses were received from 145 libraries: 111 of 211 OCLC members, 20 of 43 WLN participants, and 14 of 23 RLIN users. Approximately 10,500 requests for monographs were included. A representative sample of 807 request forms was selected for the test search. Appendix 3C shows the composition of the sample.

Each of the 807 titles was searched on each utility during February 1980. The number of holding statements was noted for each title that was found on a given data base. Those which were found but which had no holding symbols were also noted. The verification hit rates and holdings hit rates were then calculated, and the incremental benefit from linking was determined. These incremental hit rates were used to estimate the processing cost that could be avoided by an average library with access to all three data bases.

The test search results should be interpreted with some qualifications in mind:

- The test search compares the data bases as they were in February 1980, and the results thus constitute a "snapshot" of their relative standing at that time only.

- The 2-week sampling period is assumed to be typical, even though it included the Thanksgiving holiday, and several academic libraries reported that classes were not in session for part of the period.

- Almost all of the titles used in the test search had been verified either in a utility data base or in a print source by the contributing library. Some of those that were verified in a print source may have elevated the hit rate in the test search because errors had been corrected that would otherwise have prevented a hit in an online search.

Results of the Borrowing Request Test Search

Verification Hit Rates

Figure 3-1 shows the verification hit rates obtained from the test search. These are the percentages of the titles for which a record was found on each data base and on each combination of data bases.

If an OCLC user had searched for all 807 titles on that data base alone in February 1980, the hit rate would have been 77 percent. WLN users would have found records on their data base for 47 percent of the titles, and RLIN users would have found only 44 percent at that time.

Figure 3-2 compares these verification hit rates to the hit rates for current title cataloging reported in Chapter 2. The ILL hit rates are lower, because many of the ILL requests used in the test search were for material published before 1968, when the first LC-MARC records became available. Appendix 3C shows that as many as half of the requests used were for monographs published before 1968.

FIGURE 3-1. VERIFICATION HIT RATES

Results of a test search of 807 monograph requests in February 1980.

	Percentage of Titles for Which a Record was Found
One Data Base	
OCLC	77%
RLIN	44%
WLN	47%
Two Data Bases Linked	
OCLC-RLIN	79%
OCLC-WLN	79%
RLIN-WLN	53%
Three Data Bases Linked	
OCLC-RLIN-WLN	80%

FIGURE 3-2. COMPARISON OF HIT RATES FOR ILL VERIFICATION AND CURRENT TITLE CATALOGING

Data Base	Verification Hit Rate	Current Title Hit Rate (8th Week)
OCLC	77%	93%
RLIN	44%	87%
WLN	47%	87%

Incremental Verification Hit Rates

Figure 3-1 also shows the verification hit rates for combinations of the three data bases. These percentages reveal the increases in hit rate that could have been obtained by users of each utility when linked to any combination of the other data bases. The incremental hit rates are presented in Figure 3-3.

FIGURE 3-3. INCREMENTAL VERIFICATION HIT RATES FOR ALL TITLES

Results of a test search of 807 monograph requests in February 1980. Numbers represent the percentage points added to the hit rate when utilities are linked.

Host Utility	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC	--	2%	2%	3%
RLIN	35%	--	9%	36%
WLN	32%	6%	--	33%

When OCLC's resources are combined with those at either RLIN or WLN, the resulting hit rate is 79 percent. This is a small improvement for OCLC users, indicating that very few of the titles in the sample could be found on RLIN or WLN but not on OCLC. RLIN and WLN users would have increased their verification hit rates by more than 30 percentage points when linked to OCLC. If RLIN and WLN had been linked to each other but not to OCLC at the time of the test search, a user of each would have found a record for 53 percent of the titles, an increase of 6 percentage points for WLN and 9 percentage points for RLIN.

If all three data bases had been linked, the hit rate for the sample of titles would have been 80 percent. Only 3 percent of the titles were not found on OCLC but appeared on at least one of the other data bases. In other words, a three-way link would have raised an OCLC user's hit rate by 3 percentage points at the time of the test searches. If this 3 percent were stable over a whole year, and if OCLC libraries processed, say, 400,000 borrowing requests in a year, then an additional 12,000 requests could be verified online if the utilities were linked. This would be roughly 40 additional requests verified online per day, or just one per day in every 52 libraries, assuming 2,100 OCLC users.

The increase is much greater for WLN and RLIN. RLIN users could have raised their hit rate by 36 percentage points, and the increase for WLN users would have been 33 percentage points if all three utilities had been linked in early 1980. As Figure 3-3 shows, all but about 1 percentage point of the increase in both cases is attributable to the link with OCLC.

These results indicate that the benefit of linking for verification of ILL requests may be quite high for RLIN and WLN users. Libraries using either of these utilities would be able to find records for about 80 percent of their borrowing requests if they had access to OCLC's data base, compared to about 45 percent when only the host data base is searched and 53 percent when only RLIN and WLN are linked. The advantage of linking is much smaller for OCLC libraries, because the 80 percent linked hit rate represents an increase of only 3 percentage points over the hit rate when OCLC alone is used.

The hit rates reported here reflect the situation during February 1980. The present hit rates may be somewhat different, and they are likely to change even more in the future. For instance, since the test search was conducted, several research libraries have become full-scale RLIN users, and several more may join RLG in the next few months. As these libraries add their records to the data base, and as all RLG libraries perform retrospective conversions of their collections, the RLIN verification hit rate will probably improve.

More Flexible Online Searching

While OCLC users would not benefit much in the way of increased hit rate, they may receive a different type of advantage from linking. If a request has an error in either the author or title, it is sometimes not possible to find it on OCLC. To find a record for most requests, a searcher must know the first four words of a title and/or the exact main author name. For example, an OCLC title search of The Information Age: Its Development, Its Impact, edited by Donald P. Hammer, requires the following command: INF,AG,IT,D. An author search requires the entry HAM,DON,P. For an author/title search, the command is INFO,HAMM, the first four letters of the first significant word in the title and the first four letters of the author's last name.

RLIN and WLN allow any combination of title words and any portion of the author's name to be searched. It appears, then, that for RLIN and WLN libraries, the precise title is not as important as on OCLC, and this may make exact verification less important. A searcher may obtain a hit even with an incorrect author or title. Access to other utilities could increase an OCLC user's verification hit rate by providing a correct citation or another searchable field such as ISBN, LC Card Number, or government document number. Once the request is verified in this way, the OCLC library could search its host data base again to locate the requested item.

Holdings Hit Rates

Figure 3-4 shows the results of the test search for holdings hit rates. The percentages indicate the portion of titles in each category that found a record with at least one holding symbol when searched on the combination of data bases shown. (The LC symbol was counted here, but only a few of the titles hit a record with only the LC symbol, and the hit rates are not significantly affected.) The results for each data base by itself are discussed in this section, and then the linked hit rates are examined in the next section. Appendix 3D is an analysis of the requests for which no record was found.

When all 807 titles are considered, the OCLC holdings hit rate was 77 percent, and the WLN and RLIN hit rates were both 33 percent. Recall that the verification hit rates in Figure 3-1 were 77 percent for OCLC, 47 percent for WLN and 44 percent for RLIN. The difference is due to the

FIGURF 3-4. HOLDINGS HIT RATES ON COMBINATIONS OF DATA BASES

Results of a test search of 807 monograph requests in February 1980

Numbers represent the percentage of titles in each category for which a record with at least one holding was found.

	All Titles (n=807)	OCLC Titles (n=674)	RLIN Titles (n=82)	WLN Titles (n=51)	Public Titles (n=177)	Academic Titles (n=345)	Research Titles (n=285)	Sci/ Tech (n=122)	Soc/ Sci (n=184)	Bio/ Hist/ Trav (n=107)	Non-English (n=122)	All Other Subj. (n=272)	1975-1979 (n=227)	1960-1974 (n=271)	1900-1959 (n=179)	1800-1899 (n=50)*
One Data Base																
OCLC	77	76	77	87	89	81	66	71	85	83	48	86	90	77	66	64
RLIN	33	32	23	48	54	31	22	31	46	31	16	35	56	31	8	8
WLN	33	30	30	65	63	33	15	38	44	33	2	38	50	38	8	8
Two Data Bases																
OCLC-RLIN	79	78	79	89	92	82	67	74	86	85	50	87	91	79	66	66
OCLC-WLN	78	78	77	87	92	82	66	74	86	85	48	87	91	79	66	64
RLIN-WLN	43	40	34	72	73	42	26	48	45	38	16	48	67	48	11	14
Three Data Bases																
OCLC-RLIN-WLN	79	78	79	89	92	82	67	74	87	85	50	87	92	79	66	66

* 80 requests had no date.

number of titles that hit records with no holdings statements at all. All but one record found on OCLC had at least one holding symbol. On RLIN, however, 11 percent of the titles found records but no holdings, and 14 percent had no holdings listed on WLN.

The titles contributed to the test search sample by public libraries had higher holdings hit rates on each data base than those contributed by research or other academic libraries.

The hit rate for titles contributed by research libraries is three times higher on OCLC than on RLIN--66 percent for OCLC, 22 percent for RLIN, and just 15 percent for WLN. For these 285 titles, then, and at the time of the test, OCLC offered a definite advantage to research libraries in locating the requested items. The gap between OCLC and RLIN is likely to narrow, though, as the RLIN data base grows, as archive tapes are loaded, and as the RLG libraries do retrospective conversions.

The OCLC hit rate for non-English titles is three times that of RLIN. Records for about 50 percent of the non-English titles were found with holdings on OCLC, compared to only 16 percent on RLIN. WLN had holdings for just 2 percent of the non-English books. Again, OCLC had a clear advantage for research libraries at the time of the test.

The holdings hit rates by date of publication are highest for recently published works. This is to be expected, because most libraries begin their utility affiliation by adding records for newly acquired titles and do not perform retrospective conversions immediately. Much of the growth in utility membership has taken place since 1970, and consequently, a large share of the records on the three data bases are for material published in the last 10 years.

OCLC had the best performance on requests for books published before 1960. Holdings were found for about two-thirds of these requests on OCLC, while RLIN and WLN had holdings for only 8 percent.

Incremental Holdings Hit Rates

Figure 3-4 can be used to estimate the increase in holdings hit rate when the users of one utility are given access to one or both of the other data bases. This is done by subtracting the percentage shown under "One Data Base" for a given utility from the percentage listed for the desired combination. For example, in the "All Titles" column, WLN's hit rate increases from 33 percent to 79 percent when all three data bases are linked, an increment of 46 percentage points.

Figure 3-5 shows the incremental holdings hit rates for all 807 titles when each of the utilities is linked to one or both of the others. If users of WLN or RLIN were trying to locate the titles in the sample, they could have found holdings for about 46 percent of the titles only if they had access to OCLC's data base. If RLIN and WLN were linked only to each other, users of each would improve their holdings hit rate by about 10 percentage points. OCLC users stand to gain only 2 percentage points on their holdings hit rate if the three data bases are linked. This means that if an OCLC library tried to find locations for 100 monographs in a year, 77 items could be located by using OCLC alone, and only 2 more could be located by using RLIN and WLN also.

FIGURE 3-5. INCREMENTAL HOLDINGS HIT RATES

Results of a test search of 807 monograph requests in February 1980. Numbers represent the percentage points added to the hit rate when utilities are linked.

Host Utility	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC	--	2%	1%	2%
RLIN	46%	--	10%	46%
WLN	45%	10%	--	46%

Figure 3-6 presents the incremental hit rates when the sample is broken down by utility affiliation of contributing libraries. Although the subsamples are small for RLIN and WLN, these figures give a rough idea of the effect of linking on locating items that may be characteristic of those requested most often by patrons in each utility's libraries. With all three data bases linked, RLIN users could have increased their holdings hit rate by 56 percentage points. The increase for WLN was 24 percentage points, and OCLC users would have found holdings for only another 2 percent of the titles they searched.

FIGURE 3-6. INCREMENTAL HOLDINGS HIT RATES FOR REQUESTS CONTRIBUTED BY LIBRARIES FROM EACH UTILITY

Results of a test search of 807 monograph requests in February 1980. Numbers represent the percentage points added to the hit rate when the utilities are linked.

Host Utility of Contributing Library	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC (n=674)	--	2%	2%	2%
RLIN (n=82)	56%	--	11%	56%
WLN (n=51)	22%	7%	--	24%

Figure 3-7 shows the increases in holdings hit rate for titles contributed to the sample by each type of library. For each type, linking the three utilities would have added very little to OCLC's hit rate at the time of the test search. Public libraries affiliated with RLIN or WLN, however, could have increased their hit rates by 30 to 40 percentage points, and research and other academic libraries could have added about 50 percentage points if given access to the OCLC data base. In most cases, RLIN and WLN users could have had modest increases if their data bases were linked to each other only.

The incremental hit rates for non-English titles are shown in Figure 3-8. Linking would have added only 2 percent to the holdings hit

FIGURE 3-7. INCREMENTAL HOLDINGS HIT RATES FOR REQUESTS CONTRIBUTED BY THREE TYPES OF LIBRARIES

Results of a test search of 807 monograph requests in February 1980. Numbers represent the percentage points added to the hit rate when the utilities are linked.

PUBLIC LIBRARIES (n=177)

Host Utility	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC	--	3%	3%	3%
RLIN	38%	--	19%	38%
WLN	29%	10%	--	29%

ACADEMIC LIBRARIES (n=345)

Host Utility	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC	--	1%	1%	1%
RLIN	51%	--	41%	51%
WLN	49%	9%	--	49%

RESEARCH LIBRARIES (n=285)

Host Utility	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC	--	1%	0	1%
RLIN	45%	--	4%	45%
WLN	51%	11%	--	51%

rate for OCLC users trying to locate non-English books. RLIN and WLN users would have gained 34 and 48 percentage points, respectively, almost all of the increase occurring because of the link with OCLC.

FIGURE 3-8. INCREMENTAL HOLDINGS HIT RATES FOR NON-ENGLISH MONOGRAPHS

Results of a test search of 807 monograph requests in February 1980. Numbers represent the percentage points added to the hit rate when the utilities are linked.

Host Utility	Incremental hit rates when linked to:			
	OCLC	RLIN	WLN	All
OCLC	--	2%	0	2%
RLIN	34%	--	0	34%
WLN	46%	14%	--	48%

In summary, libraries using RLIN or WLN have more to gain from linking to OCLC than OCLC users would gain from a link with either or both of the other data bases. At the time of the test search, RLIN and WLN users could have added from 30 to 50 percentage points to their holdings hit rates, depending on the type of library and the type of material in question. The increases for OCLC users were never more than 3 percentage points. Although the data bases have grown in the months since the test was conducted, the basic relationship revealed by the study is undoubtedly true today and seems likely to hold for the near future.

The Economic Benefits of Linking for ILL

Online verification and location can usually be performed in less time than it would take if printed sources were used. When staff time is saved, some of the cost of processing the borrowing request is avoided. If online searching saves a substantial amount of time, then libraries may find that there is a significant economic benefit from linking. Other work could be performed in the time saved, and an ILL budget increase might not be needed to cover the normal rise in costs.

A Model for Calculating Borrowing Cost Avoided

Figures 3-9 and 3-10 present an approach to estimating the cost that is avoided when linking makes it possible to verify and locate more items online. This analysis does not attempt to project the actual cost avoidance for a given size or type of library. Some assumptions have been made to aid in explaining the method of calculation, but libraries and other interested parties should use the blank forms in Appendix 3E to make computations based on their own data and assumptions. The analysis here does indicate, however, that for most libraries the annual ILL borrowing costs avoided by linking will not be significant.

In both figures, the starting point is an assumption about the number of borrowing requests processed at an "average" library affiliated with each utility. Data on the actual average volumes were not available, and so the numbers here are estimates used for the purpose of illustration.

FIGURE 3-9. MODEL FOR CALCULATING THE VERIFICATION COSTS AVOIDED BY LINKING

	OCLC	RLIN	WLN
1. Number of monograph borrowing requests processed in a year (1)	200	500	200
2. Reduction in percentage of titles that cannot be verified online (2)	.03	.36	.33
3. Number of requests which can be verified online only if linked (Line 1 x Line 2)	6	180	66
4. Cost avoided per request if online verification is possible (3)	$2	$2	$2
5. Total cost of verification that can be avoided in a year if linked (Line 3 x Line 4)	$12	$360	$132
6. Cost avoided per request processed (Line 5 ÷ Line 1)	$.06	$.72	$.66
7. Approximate number of libraries (public, research, and other academic only)	1,430	65	59
8. Aggregate cost avoided in a year by all affiliated libraries (Line 5 x Line 7)	$17,160	$23,400	$7,788

(1) Estimate for illustration only.

(2) See Figure 3-1. Assumes all three utilities are linked.

(3) Assumes average cost of staff time is $2.50 without online searching, $.50 with online searching.

FIGURE 3-10. MODEL FOR CALCULATING THE LOCATION COSTS AVOIDED BY LINKING

	OCLC	RLIN	WLN
1. Number of monograph borrowing requests processed in a year (1)	200	500	200
2. Reduction in percentage of titles that cannot be located online (2)	.02	.46	.46
3. Number of requests which can be located online only if linked (Line 1 x Line 2)	4	230	92
4. Cost avoided per request if online location is possible (3)	$1	$1	$1
5. Total cost of location that can be avoided in a year if linked (Line 3 x Line 4)	$4	$230	$92
6. Cost avoided per request processed (Line 5 ÷ Line 1)	$.02	$.46	$.46
7. Approximate number of libraries (public, research, and other academic only)	1,430	65	59
8. Aggregate cost avoided in a year by all affiliated libraries (Line 5 x Line 7)	$5,720	$14,950	$5,428

(1) Estimate for illustration only.

(2) See Figure 3-4. Assumes all three utilities are linked.

(3) Assumes average cost of staff time is $1.50 without online searching, $.50 with online searching.

OCLC libraries vary widely in size, from research libraries that process 1,000 or more monograph borrowing requests a year to small public and academic libraries that do little or no borrowing on ILL. The average OCLC library is assumed to process 200 requests for monographs in a year. This figure is also used for WLN. The average for RLIN is higher because research libraries make up a larger portion of the clientele than at OCLC or WLN. Monograph borrowing volume is estimated to be 500 requests a year for the average RLIN user.

The results of the test search indicate that if the utilities were linked, many libraries could save processing time by verifying and locating online some requests that otherwise would have required the use of printed sources. In Figure 3-9, the increases in verification hit rate from Figure 3-3 are applied to the number of requests processed in a year to determine the number of requests that could be verified online only if the searchers had access to other data bases. Similarly, in Figure 3-10, the incremental holdings hit rates from Figure 3-5 are used to estimate the number of requests that could be located online only if the utilities were linked.

The next step is to estimate the cost that is avoided when a request is verified online instead of in a printed source. This must also be done for online location.

Studies conducted in the past 10 years indicate that the cost of processing a borrowing request in academic libraries is about $14. (Appendix 3F contains the results of a review of the literature on the unit cost of ILL.) About $10 of the $14 is direct labor cost, the expense for wages and salaries for the ILL staff who process the requests. For public libraries, the average unit cost is about $9, and the direct labor component is about $6. The $10 figure has been used here as the direct labor cost for all libraries, anticipating that inflation will soon push the average to this point.

Several studies have estimated the contribution of verification and location to the cost of processing a borrowing request when a utility is not used. Brown and McHugh (1) reported in 1976 that the time required for verification and location makes up about 34 percent of the direct labor cost. Cossar (2) found that verification alone accounted for 44 percent of the direct labor cost at one small academic library. The Westat study of ILL in academic libraries (3) did not look at the direct cost breakdown but reported that verification and location took 30 percent of the total time spent on processing borrowing requests. These findings indicate that verification and location without using an online union catalog make up between one-third and one-half of the $10 per request. If the average contribution is assumed to be 40 percent, then the direct labor cost of verifying and locating is about $4 per request.

The effort required for verification is normally greater than that for location, and it was assumed for the models that $2.50 of the $4 is spent on verification activities and $1.50 is spent on location. If these tasks can be performed online, much of this cost is avoided. Perhaps $2 of the $2.50 can be saved if a title is verified online, and $1 of the $1.50 might be avoided if printed location sources do not have to be used. Not all of the cost is avoided, of course, because some time is still spent on the online search of the host data base and the linked data base.

---

(1) M. K. Brown and A. L. McHugh, Survey of Costs in Technical Processing and Interlibrary Loan, Western Interstate Library Coordinating Organization (WILCO), Boulder, Colorado, December 1976.

(2) Bruce Cossar, "Interlibrary Loan Costs", RQ, Spring 1973, pp. 243-246.

(3) Westat, Inc., A Study of the Characteristics, Costs, and Magnitude of Interlibrary Loans in Academic Libraries, Westport, Connecticut: Greenwood Publishing Company, 1972.

To calculate the verification cost avoided per request in Figure 3-9, the $2 estimate is multiplied by the number of requests that could be verified only if all three data bases were linked. The $1 avoided for location is also multiplied by the number of requests in Figure 3-10. In each case, the result is an estimate of the cost that can be avoided in a year by the hypothetical average library when the data bases are linked.

Figure 3-11 summarizes the results of this analysis, adding the cost avoidance estimates for verification and location. An RLIN library that processes 500 monograph borrowing requests in a year could avoid nearly $600 in annual ILL costs if the utilities are linked. A WLN user that processes 200 requests in a year might save about $225. The cost avoided by most OCLC libraries would be negligible, however, because of the very small increase in hit rate for both verification and location when the data bases are combined. An OCLC member that processes 200 requests a year might avoid less than $20 of the annual cost of ILL.

Linking might enable RLIN and WLN libraries to save more than $1.00 per borrowing request, about a 7 percent reduction in the $14 average cost of processing a request. Most OCLC users would avoid only a few cents of the unit cost. The RLIN network as a whole could save more than $38,000 in annual ILL costs, OCLC libraries could avoid a total of about $23,000, and the aggregate figure for WLN is just over $13,000.

FIGURE 3-11. SUMMARY OF ILL COSTS AVOIDED BY LINKING

See derivations in Figures 3-9 and 3-10 .

Total Cost Avoided Per Library in a Year

	OCLC	RLIN	WLN
Verification Cost	$12	$360	$132
Location Cost	4	230	92
Total	$16	$590	$224

Cost Avoided Per Request

	OCLC	RLIN	WLN
Verification Cost	$.06	$ .72	$ .66
Location Cost	.02	.46	.46
Total	$.08	$1.18	$1.12

Total Cost Avoided in a Year, All Affiliated Libraries

	OCLC	RLIN	WLN
Verification Cost	$17,160	$23,400	$ 7,788
Location Cost	5,720	14,950	5,428
Total	$22,880	$38,350	$13,216

The Costs of Linking for ILL

As was noted for shared cataloging in Chapter 2, the cost avoided by linking will be partially offset by the cost of obtaining information from the other utilities. There is likely to be a fee for each search of a linked data base, and there may also be a fixed monthly or annual charge.

Other factors may also reduce the economic benefit that has been outlined above. Some of the time saved by accessing the holdings statements on other data bases will be offset by the additional time that must be spent in translating unfamiliar holding symbols. This may only be a temporary problem while searchers are becoming familiar with the new symbols.

If the holdings of libraries using one utility are made known to libraries using other utilities, then some libraries may receive more lending requests than they did before linking. This may be true particularly for those that have done retrospective conversion of their collections. Some libraries may assume that another library is a likely source because of the size of its collection. The cost of processing the added lending requests in these cases would partially replace the cost avoided in processing the borrowing requests.

Summary

The calculations in this section indicate that the economic benefit of linking for ILL is much smaller than that of linking for shared cataloging. Many libraries would avoid less ILL cost than estimated in Figures 3-9 and 3-10 because they simply do not receive very many borrowing requests from their patrons in a year. However, the primary benefit of linking for ILL does not lie in the cost avoidance. The opportunity to verify a title and locate an item on other data bases, when combined with ILL policies that encourage cooperation, may improve the fill rate at many libraries, and this enhancement of patron services could be an important justification for linking.

CHAPTER 4

THE BENEFITS OF LINKING FOR REFERENCE SEARCHING

Introduction

The question of the potential benefits of linking for reference searching differs from the related issues regarding shared cataloging and ILL. Reference searching may involve library patrons as well as staff, and a key issue that must be addressed is the extent to which the bibliographic utilities could accommodate the anticipated traffic from public access to online catalogs at any point in time.

With this in mind, the study attempted to resolve two issues with respect to reference searching. First, estimates were made of the number of terminals and central processors that would be needed to support reference searching if local card catalogs were immediately closed and catalog searches used only online union catalogs. The second issue is the expected demand for searching more than one online union catalog, assuming that local catalogs are not closed.

Resources Required to Replace Local Catalogs with Online Union Catalogs

The feasibility of providing online catalog searching in place of local catalogs was examined by estimating: (1) the catalog traffic in libraries affiliated with a bibliographic utility, (2) the number of terminals that would be needed to support this traffic at peak hours, and (3) the computer resources that would be required to perform the data base searches. Details of the analysis and supporting data are contained in Appendix 4A. The findings may be summarized as follows:

- Peak catalog traffic in 1,227 public and academic libraries is 30,000 patrons between 3:00 and 4:00 p.m. EST. (See Figure 4A-10 in Appendix 4A.)

- The number of computer terminals needed at the peak hour was estimated to be 17,780, assuming a 4-minute service time per patron and a 90 percent probability that no patron would wait in line longer than 2 minutes.

- Approximately 95 computer-hours per day are required to support online catalog searching with the present system configurations. OCLC would require about 92 of these hours, roughly the equivalent of 5 more central processors working 18 hours a day.

These estimates suggest that if all searches at local catalogs are ultimately to be performed on online union catalogs, considerable additional facilities will be required, especially at OCLC. Other obstacles include the cost of telecommunications, the deterioration of response time with high volumes, the requirement of simplicity in user interfaces, and the need for high reliability so that catalog searches are not interrupted or prevented by a failure in the system. The conversion to total online searching is not likely to occur in the near future, and for this reason the remainder of this chapter concentrates on how linking the utilities can affect reference searching as it is currently conducted.

The Benefit of Linked Reference Searching

Public or staff use of an online union catalog extends the opportunities for reference searching far beyond the local catalog. Linking the bibliographic data bases at the utilities would give a searcher access to still other resources that might aid in a reference search. Some cost may be avoided because staff time is saved in online searching, but the primary advantage of linking is that library service is enhanced when patrons and staff can locate bibliographic records not available from the host utility or, as is the case for OCLC libraries today, when they can use searching capabilities on other data bases that are not provided by the host utility. If the demand for linked reference searching is generally high, then individual libraries may find that their reference searching requirements alone provide a strong justification for linking.

In this study, the benefit of linked reference searching was equated with the expected number of "successful" searches of a second bibliographic data base. A successful search is defined to mean a catalog search in which some useful information is found. (The searcher may not consider a successful search to be a completed search, though, and may try to expand the results by searching elsewhere.) In general, then, the more linked searches that result in useful information, the greater the benefit to the searcher and the better the service by the library.

The benefit of linked reference searching may therefore be projected in the following manner: (1) estimate the number of reference searches that an average library would conduct online using its host utility, (2) estimate the portion of these searches that would be repeated on a second data base, and (3) estimate the portion of the linked searches that would be successful.

There are several reasons for a patron or staff member to conduct an online reference search. For example, the searcher may wish to expand or verify the results of a local catalog search; the local catalog search may not result in any useful citations, and the online data base may be a convenient next step in the search; and some people may simply choose to use the online catalog directly, without consulting the local catalog first. No data could be found on the percentage of reference searchers that would use a terminal first if given the opportunity, nor are there any studies that report on the percentage of searches at the local catalog that are considered incomplete by the searcher and might thus be pursued further in an online search. Because of this difficulty in estimating the potential level of online reference searching, it was not possible to estimate the total benefit from linked reference searching.

Previous studies do provide a basis for estimating the number of searches that "fail" at the local catalog--those that result in no useful information at all--and it was therefore possible to estimate the demand for linked reference searching that is attributable to these failed searches. The figures derived in this chapter thus represent only a portion of the potential benefit from linked reference searching, and the size of that portion is not known.

A Model for Estimating the Volume of Linked Reference Searching

Figure 4-1 presents a model for estimating the benefit of linked reference searching at an average public, academic, and research library using each utility. Each step in this method is discussed below, and the sources of the assumptions and the data underlying the calculations are explained. The methodology is presented in this way so that interested libraries and others may alter the assumptions or substitute data for certain items and then assess the impact of these changes on the final estimates. A blank version of the model is provided in Appendix 4B.

Line 1, the average daily circulation for each type of library affiliated with each utility, was used to estimate catalog usage. No data are available on the volume of local catalog usage in the libraries using each of the three utilities. However, as explained below, previous studies have indicated that catalog usage and circulation are related, suggesting that if the ratio between catalog use and circulation is known for a given type of library, then one may be used to predict the other. Since circulation data is generally available, the catalog usage was estimated in this way.

For each public, academic, and research library using RLIN and WLN, annual circulation figures were obtained from published sources of statistics or from inquiries to selected libraries. Special libraries were not included in this study. The large number of OCLC members made it necessary to collect circulation data from a representative sample of libraries. A random sample of the OCLC public and academic libraries in the

FIGURE 4-1. A MODEL FOR ESTIMATING THE BENEFIT OF LINKED REFERENCE SEARCHING

	OCLC			RLIN			WLN		
	Public	Academic	Research	Public	Academic	Research	Public	Academic	Research
1. Average daily circulation	1500	350	2000	7500	850	2500	1900	250	2200
2. Number of items circulated per use of the catalog	8	1	1	8	1	1	8	1	1
3. Number of people using the catalog each day (1 ÷ 2)	200	350	2000	940	850	2500	240	250	2200
4. Number of searches by each person using the catalog	1.24	1.32	1.32	1.24	1.32	1.32	1.24	1.32	1.32
5. Number of catalog searches each day (3x4)	248	462	2640	1166	1122	3300	298	330	2904
6. Percentage of searches that fail at the local catalog	.30	.23	.23	.30	.23	.23	.30	.23	.23
7. Number of searches that fail each day (5x6)	74	106	607	350	258	759	89	76	668
8. Percentage of failed catalog searches that are searched on-line	.10	.50	.70	.10	.50	.70	.10	.50	.70
9. Number of failed searches that are searched on-line each day (7x8)	7	53	425	35	129	531	9	38	468
10. Percentage of failed searches on the home data base	.70	.70	.70	.50	.50	.50	.50	.50	.50
11. Number of failed searches on the home data base each day (9x10)	5	37	298	18	65	266	5	19	234
12. Percentage of failed searches that are searched on a second data base	.30	.50	.60	.50	.50	.50	.50	.50	.50
13. Number of failed searches that are searched on a second data base each day (11x12)	2	19	179	9	33	133	3	10	117
14. Percentage of successful searches of the second data base	.20	.20	.20	.30	.30	.30	.35	.35	.35
15. Number of successful searches of the second data base each day (13x14)	--	4	36	3	10	40	1	4	41

NOTE: These calculations do not project the full level of linked reference searching. The demand for linked searching is derived from the number of searches that "fail" at the local catalog---searches that do not result in any useful bibliographic information. A comprehensive analysis should also consider other sources of online searching for which data are not presently available, including (1) searches that do not fail at the local catalog but which are pursued further in order to expand or verify the results, and (2) searches by people who use the terminal directly, without looking in the local catalog first.

A blank copy of this form may be found in Appendix 4B.

U.S. was selected. Annual circulation data was obtained for these libraries in the same way that it was obtained for the other two utilities, and total circulation for the entire OCLC network was estimated.

Average annual circulation was calculated for each of the three library types by dividing total circulation in each category by the number of libraries in that category. This was done for each utility. Since most public libraries are open 6 days a week, the total annual circulation for public libraries was divided by 313 to obtain an estimate of daily circulation. Academic and research library annual circulations were divided by 360, since most are open 7 days a week but closed on holidays. The results are shown on Line 1 of Figure 4-1.

The estimated daily circulation for research libraries is roughly the same for all three utilities. OCLC and WLN academic libraries are also close. Academic libraries that use RLIN have a higher daily average, though, because 8 of the 11 institutions counted in this category were large academic libraries. The same is true for public library circulation. The WLN average is a little higher than OCLC's, but the figure for RLIN is huge in comparison because the 12 RLIN public libraries counted included some very large systems, notably the Los Angeles County Public Library.

These estimates were based on 1977 circulation data from the National Center for Education Statistics and the American Library Directory, the latest information available. No attempt was made to adjust them to a 1980 level, because a reliable estimate of the annual circulation growth rate was not available.

Local Catalog Traffic

Line 2 of Figure 4-1, the number of items circulated per use of the catalog, estimates the ratio between items circulated and catalog uses. A catalog use study by Ben-Ami Lipetz (1) found that the number of people using Yale University's main catalog for a given year was virtually the same as the number of books circulated in that year. In other words, the ratio of catalog usage to circulation was 1:1. James Krikelas confirmed this relationship in another study published shortly after Lipetz's work (2). Since the Lipetz study is recognized as a methodological standard, the 1:1 ratio was used to typify the catalog activity in all academic libraries. This was an over-simplification, but no data were available to indicate how the ratio might vary among academic libraries.

There are apparently no studies to suggest a ratio of circulation to catalog use for public libraries. This ratio is likely to be quite different from 1:1 though, primarily because public library patrons generally browse for recreational reading material more than academic patrons, and thus they generate more circulation activity per catalog use.

---

(1) Ben-Ami Lipetz, "Catalog Use in a Large Research Library," Library Quarterly, 42:1 (January 1972), pp. 133-134.

(2) James Krikelas, "Catalog Use Studies and Their Implications", in Melvin Voigt, ed., Advances in Librarianship, Vol. 3, New York: Seminar Press, 1972.

The project staff made direct observations of catalog use at the main public library and one typical branch in a medium-sized metropolitan area. The catalog usage data were then normalized by a factor based on a composite curve for catalog traffic in public libraries. This curve was derived from a study by Butler et al (3) of catalog use at the Los Angeles Public Library. The normalized observations were then compared to circulation data, and the ratio was estimated to be about 1:8. That is, for every catalog use in a public library, 8 items are circulated.

When Line 1 is divided by Line 2, the result is an estimate (Line 3) of the number of people who use the catalog each day in the libraries of the respective utility memberships.

Line 4 contains an estimate of the number of searches by each person using the catalog. Line 5, the number of catalog searches each day, is the product of Lines 3 and 4.

Ruth Hafter (4) recently reviewed the catalog search literature and tabulated the findings of a number of studies on the types of searches that are performed. For 11 academic library studies listed in the review, the weighted average percentage of known-item searches was 66 percent. Subject searches plus other types of unknown-item searches thus comprise about 34 percent of all catalog searches in academic libraries. The public library average from six studies was 42 percent known-item and 58 percent unknown-item searches.

---

(3) Brett Butler, Martha W. West, and Brian Aveney, "Library and Patron Response to the COM Catalog: Use and Evaluation, A Report of a Field Study of the Los Angeles County Public Library System", Information Access Corporation, Los Altos, California, 1979.

(4) Ruth Hafter, "Type of Search by Type of Library", Information Processing and Management, 15 (1979), pp. 261-264.

A study by Renata Tagliacozzo and M. Kochen (5) suggests that patrons in public libraries perform an average of 1.27 subject searches per trip to the catalog. For academic libraries, the rate was 1.52 subject searches per patron. Tagliacozzo and Kochen also found that the rate of known-item searches was 1.21 in both cases.

The results of these studies may be combined to yield an estimate of the number of searches performed by each person using the local catalog. For public libraries, 58 percent of the people conduct 1.27 unknown-item searches and 42 percent conduct 1.21 known-item searches per trip to the catalog. The weighted average is about 1.24 searches per person. For academic libraries, 34 percent of the people conduct 1.52 unknown-item searches and the other 66 percent conduct 1.21 known-item searches. The weighted average in this case is 1.32 searches per person.

Reference Searches of the Host Utility's Data Base

Line 6 is the percentage of searches that "fail" at the local catalog. F. W. Lancaster (6) has reported on the success and failure rates of catalog searches. The data from that study indicate that the failure rate for known-item searches is about 40 percent in public libraries and about 30 percent in academic libraries. Subject searches fail about 15 percent of the time in public libraries and about 20 percent of the time in academic libraries. When these failure rates are weighted by the porportions reported by Hafter for each type of search, the average failure rate for all searches is about 30 percent for public libraries and about 23 percent for academic libraries. Hence the product of Line 5 and Line 6 is the number of searches that fail each day at the local catalogs of utility member libraries (Line 7). This is the number of local searches that are candidates for an online search.

---

(5) Renata Tagliacozzo and M. Kochen, "Information-Seeking Behavior of Catalog Users", Information Storage and Retrieval, 6:5 (December 1970), pp. 363-381.

(6) F. W. Lancaster, The Measurement and Evaluation of Library Services, New York: Information Resources Press, 1977, Chapter 2.

Line 8 contains an estimate of the percentage of failed catalog searches that are searched online, and Line 9 is the resulting number of failed catalog searches that are searched online each day.

Many candidates for online reference searching will not continue beyond the local catalog. Some patrons will not wish to try to obtain a book on interlibrary loan; others will give up because they do not believe that even the online union catalog will have what they want. Some may simply not be aware of the online reference searching service.

The assumption was made that public library catalog users would pursue the search further only one time in ten. The "perseverance rate" will probably be higher at academic libraries than at public libraries because of the importance some faculty and students place on obtaining at least a citation for a book on a subject they are researching. Others will simply insist on the most complete search possible. It was assumed that the patrons and staff of most academic libraries will conduct online searches for half of the searches that fail at the local catalog. At research libraries, the percentage is likely to be even higher, and 70 percent was used here. It must be emphasized that these are estimates only and the actual percentages will vary considerably from one institution to another.

Line 10 is an estimate of the percentage of failed searches of the host data base, and Line 11 is the number of failed searches on the host utility each day.

OCLC members cannot presently perform subject heading or title keyword searches, and so their failure rate for subject searching is by default 100 percent. Since roughly half of all catalog searches are subject searches in public and academic libraries combined, the overall failure rate on OCLC is at least 50 percent. These searches are "failures" because they cannot be conducted. Failed author searches and known-item searches probably raise the overall failure rate to 70 percent.

It is equally difficult to estimate the failure rates of reference searches at the other utilities. It was assumed for this analysis that the failure rate for both RLG and WLN members is 50 percent. This is somewhat lower than the estimate used for OCLC because of the subject search capability available with RLIN and WLN.

The Volume of Linked Reference Searching

Line 12 is the percentage of failed searches that might be expected to be repeated on another utility's data base. About half of the

RLIN and WLN failures are assumed to be pursued further via a link to another utility. This figure is higher than the 10 percent perseverance rate that was estimated for failed catalog searches because it seems reasonable to believe that many of the people who try an online search once will be willing to search again if they fail. Since OCLC does not support subject searching, a linked data base would be the first resort for those who wish to conduct online subject searches. The perseverance rate used on Line 8 was applied to roughly half of the searches. The 50 percent used for RLIN and WLN on Line 12 was applied to the other half of the OCLC searches.

Line 13 is the product of Line 11 and Line 12 and represents the number of failed searches on the host data base that would be searched on a second data base. These numbers are estimates of the daily volume of linked reference searching by the users of each utility. OCLC research library patrons and staff might conduct about 180 searches a day per library on RLIN and WLN, taking advantage of the subject searching capability at these utilities. The volume at RLG and WLN research libraries would be about 120 or 130 a day, lower than the OCLC figure because many searches can be satisfied by the host data base. Searchers at other academic libraries would use a linked data base from 18 to 30 times a day at each library, while linked searches might be conducted from 2 to 10 times a day at an average public library.

Figure 4-2 shows how the volume estimates on Line 13 of Figure 4-1 may be converted to daily totals for each utility. OCLC's public, academic, and research library users would conduct more than 33,000 linked reference searches a day this year if given access to the other data bases. RLIN

FIGURE 4-2. DERIVATION OF TOTAL LINKED
REFERENCE SEARCHES PER DAY

OCLC

Type of library	Number of libraries(1)	Linked reference searches per day per library(2)	Total linked reference searches per day
Public	287	2	574
Academic	1,072	19	20,368
Research	71	179	12,709
Total	1,430		33,651

RLIN

Type of library	Number of libraries(1)	Linked reference searches per day per library(2)	Total linked reference searches per day
Public	17	9	153
Academic	29	33	957
Research	19	133	2,527
Total	65		3,637

WLN

Type of library	Number of libraries(1)	Linked reference searches per day per library(2)	Total linked reference searches per day
Public	21	3	63
Academic	36	10	360
Research	2	117	234
Total	59		657

(1) Public, academic and research libraries only. OCLC count is for May 1980; RLIN and WLN counts are for June 1980.

(2) The demand for linked searching is derived from the number of searches that "fail" at the local catalog---searches that do not result in any useful bibliographic information. A comprehensive analysis should also consider other sources of online searching for which data are not presently available, including (1) searches that do not fail at the local catalog but which are pursued further in order to expand or verify the results, and (2) searches by people who use the terminal directly, without looking in the local catalog first.

libraries would use a link for about 3,600 searches a day, and WLN users would conduct 650 searches a day. It must be stressed that these figures are estimates of the demand for linked reference searching that would result only from failures at the local catalog. The total daily search load is likely to be significantly greater.

Successful Linked Searches

Line 14 is the assumed percentage of successful searches of the second data base, and Line 15 the resulting number of successful searches of the second data base each day.

Since OCLC's data base is not subject indexed, patrons or staff of RLIN and WLN libraries will usually fail to find what they are looking for when they conduct a subject search on a second data base. RLIN libraries may search only WLN, and this smaller data base would not add significantly to the subject-indexed resources on RLIN. WLN users might find that subject access to RLIN's data base is more useful because of its different composition, but the advantage would still be small.

The high OCLC hit rates reported in Chapters 2 and 3 indicate that known-item searches and author searches on OCLC have an excellent chance of succeeding, though. Altogether, perhaps 10 percent of RLIN's linked subject searches and 50 percent of the other linked searches would be successful, for a composite success rate of 30 percent. For WLN libraries, it is assumed that 20 percent of the linked subject searches would succeed, along with 50 percent of the others, giving a composite rate of 35 percent. OCLC users might be expected to have the highest success rate for linked subject searching, say 30 percent, and a quite low success rate for other linked searches because so much can be found on the OCLC data base. Perhaps 10 percent of the known-item and author searches would succeed on RLIN or WLN after failing on OCLC, and this gives OCLC libraries a composite success rate of 20 percent.

Line 15 is the number of successful linked reference searches that would be performed each day by libraries of each type. These figures represent an average benefit from linking for reference searching, based on the assumptions and data described here. Individual libraries may substitute their own assumptions and use the blank worksheet in Appendix 4B to derive estimates that reflect their own circumstances more accurately.

Conclusions

It appears that library reference services can benefit substantially from linking. Patrons or staff in research libraries would average 30 to 40 successful searches in a day on a second data base. In other academic libraries, 5 to 10 successful searches would be conducted each day. Public libraries would be successful on 1 to 3 searches a day. These figures add up to over 2 million successful reference searches a year--that is, more than 2 million instances of a second data base providing useful information when no suitable citations were available from either the local catalog or the host utility.

The projections developed here indicate that the aid to reference searching alone may be enought to justify linking. It must be re-emphasized that these estimates reflect only a portion of the potential advantage from linking. Only searches that "failed" at the local catalog have been counted as candidates for online reference searching, and thus for linked reference searching. A comprehensive analysis would also have to consider (1) online searches for the purpose of expanding or verifying the results of a catalog search, and (2) online searches attempted without consulting the manual catalog first. The approach used in Figure 4-1 may be used to produce demand estimates for these categories of reference searches when data become available. For now, it can only be said that the true benefit of linked reference searching comprises many more successful searches a day than the figures presented here indicate, perhaps totaling several million additional successful searches a year.

CHAPTER 5

DERIVATION AND PRELIMINARY ANALYSIS OF LINK ALTERNATIVES

Introduction

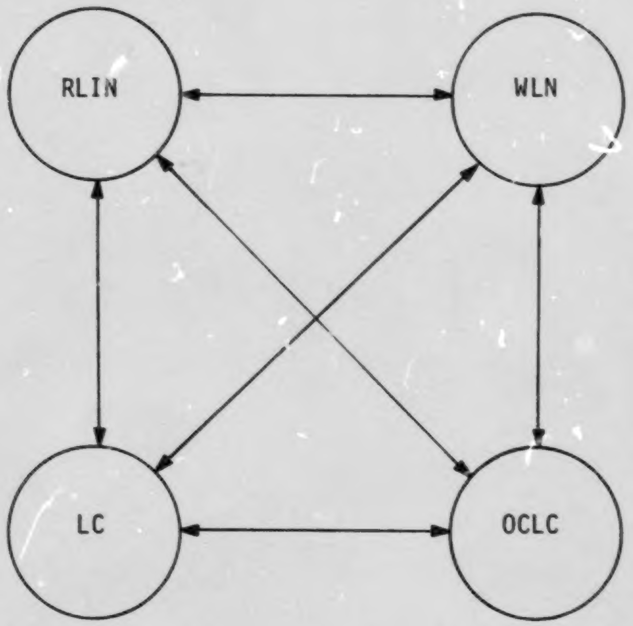
The findings of Chapters 2, 3, and 4 suggest that most libraries now using a utility could benefit in one or more service areas if given access to the unique resources and capabilities of the other utilities. When all four data bases are linked, the potential hit rate for current titles is 96 percent, enabling some libraries to avoid original cataloging for many of the current titles they acquire. In addition, RLIN and WLN users could perform online verification and location for many more borrowing requests if they had access to the OCLC data base. The probable demand for linked reference searching at many libraries provides another impetus for linking.

Figure 5-1 depicts the network configuration suggested by these findings: all four institutions linked to each other. This chapter and the next now address the alternative ways in which the four data bases could be linked.

Many linking configurations were analyzed to determine which one appears to be the most practical to implement and operate. A key consideration throughout this analysis was the balance between the benefits to the user community and the cost of the link needed to provide those benefits.

This chapter describes the methodology used to define the link alternatives and provides the analytical framework used to select some alternatives for further consideration. Three of the alternatives are the subject of a more detailed evaluation in Chapter 6.

FIGURE 5-1. NETWORK CONFIGURATION FOR LINKING



Several assumptions were made in order to limit the array of conceivable alternatives to those that could be effectively analyzed within the scope of this study. First, it was assumed that any linking configuration tying the utilities together would be homogeneous; that is, the six legs of the four-way network shown in Figure 5-1 would be of the same type. This was done for the sake of simplicity and to enable comparison of the alternatives on an equal basis.

Second, primary consideration was given to those link alternatives that can be implemented within the next 3 or 4 years. Networking and data base technologies are evolving very rapidly, and there are a number of ways in which the storage and access of bibliographic data could change within the next decade. For example, distributed data base systems could allow the utilities to decentralize their resources, perhaps enabling them to improve and expand their services. Increased local automation at libraries, including the use of more "intelligent" terminals and minicomputers, may make it possible for the utility user to switch at will between sources of information. By the 1990's, home delivery of information services via cable TV or small, inexpensive terminals may even change the way that libraries are used. Some of these developments are technologically feasible today and could play an important future role in nationwide linking. However, use of such techniques in the initial linking phases would require radical changes by all utilities, and opposition to such approaches could be overwhelming. This study examined link alternatives that use existing terminals and utility services, arrangements that could provide the anticipated benefits of the link while requiring minimal changes in existing operations.

What is a Link Alternative?

Any link alternative has several features. The four principal characteristics are:

(1) Exchange Method. The exchange method is the way in which an institution provides another institution with access to its resources. The three exchange methods considered in this study are (a) delivery

of computer tapes, (b) batch data transmission of records, and (c) online access.

(2) Data Exchanged. This is the portion of a utility's data base to which a linked user has access. In other words, the user could have access to all of a utility's data and services or to only a portion of them. The user might also be able to initiate requests for specific records from the linked data base.

(3) Communication Method. The communication method is the physical means by which data are transmitted between utilities. If the exchange method is tape delivery, how are the tapes sent? For the other methods of exchange, what kind of telecommunication lines can be used?

(4) User Interface. The user interface is the manner in which the terminal user becomes involved in obtaining data from a linked utility. The choice of user interface is an important aspect of linking because it determines the extent to which the user must learn the "language" of the linked utility.

Other characteristics might also be considered. For example, locus of management control of the linked network could be examined. Is each utility responsible for the administrative work associated with the operation of a link, or should there be an independent organization that manages the linked network? Direction is another characteristic. A link can operate in both directions between two participants, or the connection may go only one way, with one of the institutions acting as a passive supplier to the other. It is assumed here that all of the links in an OCLC-RLIN-WLN-LC network would be two-way exchanges.

As shown in Figure 5-2, each characteristic in turn consists of a number of options. These options were selected because they represent the most likely components of links that could be established in the next few years. Link alternatives may be "assembled" by selecting one option from those listed for each characteristic. For example:

Exchange Method: Direct tape delivery

Data Exchanged: All newly created member-generated records

Communication Method: U.S. Postal Service

FIGURE 5-2. OPTIONS WITHIN EACH CHARACTERISTIC OF A LINK ALTERNATIVE

CHARACTERISTIC 1: EXCHANGE METHOD

Direct Tape Delivery	A magnetic tape containing search requests is sent to the utilities to be searched. Responses are put on one or more tapes and sent back to the requesting institution, where they are made available to the user on-line.
Tape Delivery With Clearinghouse	Each utility sends its search request tapes to an organization that eliminates duplicates, repackages the data, and sends the new tapes to each destination utility. Responses are reformated by the clearinghouse and returned to the requesting utilities in the same way.
Direct Batch Data Transmission	Requests are saved on tape or disk and periodically are transmitted to the destination utilities over a communication line. Responses are returned to the host utility in the same manner.
Batch Transmission With Clearinghouse	Requests are saved on tape or disk and periodically are transmitted over a communication line to an organization that eliminates duplicates, repackages, and transmits the data to each destination utility. Responses are reformated by the clearinghouse and transmitted to the appropriate utilities.
On-Line Access	The user interacts in real time with the data base at the linked utility.

CHARACTERISTIC 2: DATA EXCHANGED

All Non-LC-MARC Records	All records that have been created by a member library are converted to a standard format and sent to other utilities.
Selected Non-LC-MARC Records	Member generated records that are thought to be useful to users at another utility are converted to a standard format and sent to that utility.
All New Non-LC-MARC Records	All newly created records are periodically converted to a standard format and sent to the other utilities.
Selected New Non-LC-MARC Records	Newly created records that are thought to be of greatest value to others are periodically converted to a standard format and sent to the other utilities.
Requested Records Only	Users who are not satisfied with the results of a search on their host data base request to repeat the search on one or more other data bases. Only the records found in the linked data base are sent back to the requesting utility.

FIGURE 5-2. Continued

CHARACTERISTIC 3: COMMUNICATION METHOD

Tape Delivery Only

U. S. Postal Service	First class delivery time is 3 to 5 days.
Parcel Service	United Parcel Service and others have delivery time of 3 to 5 days.
Commercial Courier	Federal Express, Emery, etc. deliver in one day for $15 to $25 per tape.

Batch Transmission and On-Line Access

Value-Added Network (VAN)	VAN operators lease common carrier lines and give the user the added "value" of services like packet-switching, error correction, and message storage. Tymnet and Telenet are the largest VAN's.
Direct Dial Network	AT&T's Direct Distance Dialing facilities are used as needed.
Leased Dedicated Lines	Each pair of utilities is connected by a private line leased from AT&T.

On-Line Access Only

Private Packet-Switched Network	Privately-owned nodal processors switch "packets" of data over a network of leased lines.
OCLC Network	OCLC's multidrop leased lines are used to connect all participants.

CHARACTERISTIC 4: USER INTERFACE

Numeric Keys Only	Linked data bases may be searched only by LC Card Number, ISBN, or any other numeric key that at least two utilities have in common. The host system adds the appropriate search prefix automatically.
Native Mode	The host utility's computer effectively converts the user's terminal into a part of the other utility's network. The user then searches and receives responses at the terminal in the "language" of the linked utility.
Translation Mode	Search requests are automatically translated into a standard format, sometimes with the aid of "fill-in-the-blanks" information provided by the user at the terminal. Responses from the linked utility are converted back into the format used by the host utility.

Some alternatives also involve a user interface option. For example:

Exchange Method: Online access

Data Exchanged: Full access to each linked data base

Communication Method: Value-added network (VAN)

User Interface: Full translation of numeric keys only

Preliminary Analysis of Link Alternatives

A large number of link alternatives can be identified using the approach outlined above, but many of these were rejected early. The full list of alternatives was reduced by eliminating the options in Figure 5-2 that were judged unsuitable because of technical or economic drawbacks. The rationale for eliminating some of the options is explained in the following sections.

Exchange Method Options

Direct tape delivery, direct batch transmission, and online access options were selected as the most promising exchange methods. The clearinghouse options were dropped because even though economies of scale might be obtained by sorting and consolidating requests and responses at a single location, the issues of management structure, administrative procedures, and security measures were felt to be beyond the scope of this study.

Data Exchanged Options

The transfer of requested records was chosen as the most desirable type of data exchange. In this option, if the user is not satisfied with results of a search of the host data base, the search may be repeated on a linked data base.

Replication of all member-generated records on each utility is not economical. This form of data exchange would require each utility to maintain a data base of 7 or 8 million records. The additional storage and processing costs would be unreasonable economic burdens.

An alternative to exchanging all member-generated records is to exchange only those records that are considered likely to be used. Data from OCLC, summarized in Figure 5-3, suggest that a relatively small number of OCLC's records have received most of the usage. Three definitions of "likely to be used" are given, and the number of OCLC records that would be included is shown in each case. After duplicates are eliminated, it may be that RLIN and WLN would have to replicate less than 5 percent of OCLC's member-generated records (about 200,000 records today) in order for their users to get most of the benefit possible from access to OCLC's data. (This assumes that many RLIN and WLN users catalog the same material that OCLC users have cataloged.)

FIGURE 5-3. USAGE OF MEMBER-GENERATED RECORDS AT OCLC

Total Member-Generated Records	2,839,000
Member-Generated Records That Have Been Used at Least Once	1,533,000

Definition of "Likely to Be Used"	Number of Member-Generated Records	Percent of All Member-Generated Records	Percent of Used Member-Generated Records
More than 10 libraries have used	312,000	11%	20%
More than 20 libraries have used	138,000	5%	9%
More than 50 libraries have used	24,000	1%	2%

Source: OCLC Cataloging Data File Statistical Report, November 30, 1978.

The one-time exchange of only the most-used records thus appears to be a promising link alternative. However, this option was excluded for two reasons. First, statistics similar to those quoted in Figure 5-3 for OCLC were not available from RLG and WLN. Conclusions about the feasibility of linking in this manner should await the closer examination that such data would support. Second, a one-time exchange would ignore all records entered into each data base after the exchange, at least until another historical usage analysis could determine which records should become part of a new exchange. These problems make a one-time replication of existing records less attractive than alternatives that involve an ongoing exchange of data.

The exchange of all newly created records was also eliminated from the list of options. The data presented in Chapter 2 indicate that the very large number of records that would be exchanged each year would produce relatively little benefit in the way of original cataloging avoided. OCLC users would benefit from obtaining about 6,000 records from RLIN this year (about 600,000 original cataloging records added this year times a 1 percent increase in hit rate if OCLC is linked to RLIN). In the same year, RLIN will produce approximately 100,000 new records through original cataloging efforts, and at least that many more from retrospective conversion activity. This means that OCLC would receive perhaps 200,000 newly created records from RLIN, while only 6,000 would actually be needed by OCLC users cataloging newly acquired material. On the other end, RLIN would receive OCLC's 1,000,000 new member-generated records (600,000 of which are original cataloging records) but would only be able to use about 7,000 records in cataloging newly acquired material (100,000 original cataloging records added at RLIN this year times the 7 percent increase in hit rate if linked to OCLC).

While these calculations are imprecise, they suggest that exchange of all newly created bibliographic records would be an inefficient method of obtaining the desired benefits from linking. A huge number of records would be exchanged and stored at the four institutions each year, and yet relatively few would be used to avoid original cataloging. The cost of storing the million records received each year from OCLC makes this option unsound for RLIN and WLN.

The possibility of exchanging only selected new member-generated records was also considered. However, there are insufficient data to determine which of the newly created records might be useful to a cataloger using another data base. Moreover, the selection criteria would need to be carefully developed taking into account the multiple needs and preferences of different types of libraries. These issues were not addressed in detail in this study.

Communication Method Options

The communication options selected for further consideration are: (1) mail, for the tape delivery alternatives, (2) direct dial-up, for the batch transmission alternatives, and (3) dedicated leased lines, for the online alternatives.

It is convenient to ignore the differences between mail, parcel service, and courier as communication methods for tape delivery. These options differ in delivery time and reliability, but the difference in cost is negligible when compared to the other costs of linking. Tape delivery is considered here to have just one communication method: "mail".

It is assumed that the four institutions will not jointly develop their own *packet-switched network*. The volume of linked searching, discussed in Chapter 6, would not justify the great expense for designing, implementing, and operating a private switched network.

The use of a *value-added network (VAN)* such as Telenet or Tymnet has some advantages. OCLC and RLIN both use a VAN now and already have the interface processors in place, although additional facilities would be needed for a host-to-host connection. The most attractive alternative to using a VAN is to *lease a long-distance telephone line* and dedicate it to the link traffic.

The development of transmission protocols for linking may be easier if a VAN is used. However, these commercial networks have not been used extensively for host-to-host links. VAN's have been used mainly to carry messages between a terminal and a computer, and they currently handle computer-to-computer communication for only a few limited applications. An interface protocol conforming to the X.25 standard is required for a synchronous network interface, the preferred form of host-to-host connection. Of the three types of front-end processor used at the utilities--COMTEN at WLN and LC, TANDEM at OCLC, and PDP 11/60 at RLIN--only the TANDEM now offers an X.25 protocol. COMTEN representatives state that a similar interface will be available for their communications processors by late 1980.

Linking via a VAN would be less expensive for each utility than using leased lines. The cost of a 4,800 baud synchronous network interface with a VAN is between $750 and $1,000 a month for each participant. The leased line charges vary with distance, and in a four-way link in which each pair divides the line costs, the total cost per utility would be about $2,000 per month today. The reliability of both approaches is good, with VAN's holding a slight edge because of the backup facilities that are routinely provided. The VAN approach might also be easier to manage, since using leased lines would require sharing the monthly costs and coordinating the installation and line monitoring.

In summary, both VAN's and leased lines are feasible communication methods. However, the major problems with the VAN approach are the lack of an X.25 protocol for the communications processors used by RLIN, WLN, and LC and the general lack of experience with the use of VAN's for host-to-host rather than terminal-to-host links. The VAN option was eliminated from further consideration because of these problems, but this communication method should be examined again carefully when the utilities begin the detailed design for linking.

Another communication option initially considered was the use of OCLC's telecommunications network as the basis for the linked network. In this case, the computers used by RLIN, WLN, and LC would in effect become concentrators on OCLC's network and could access OCLC's data base in this manner. Each host would require interface software to make it appear as a device compatible with the OCLC network. OCLC member libraries could be switched over the network, either by their regional broker or by an OCLC computer, and could access data at the other institutions. While this is an intriguing option, there has not been sufficient study of the effect this arrangement might have on the performance of the OCLC network in serving its own users.

An online link between the computers of any two utilities would have to be continuously available for several hours a day in order to accommodate users in different time zones and to avoid bunching the link traffic into too short a period. The cost of maintaining a dial-up computer-to-computer connection via the public telephone network for, say, 10 hours a day, would be prohibitively high. Therefore, the dial-up option was also excluded for online link alternatives.

Similar reasoning argues against the use of a dedicated leased line for batch transmissions. The data transfer would probably occur in a single "burst" once a week. Even if the transmission were to be made once a day, the expense for a dedicated line would not be justified.

User Interface Options

The native mode and translation mode interface options were selected for further consideration.

The use of numeric keys only would be the easiest translation to implement because the numbers (LCCN and ISBN might be used) are handled in a similar manner by all of the systems. Either the host utility or the linked utility could easily add the appropriate search command and suffix. However, restricting linked searching to these numeric arguments would severely limit the benefit that might be obtained from linking. Few interlibrary loan requests contain either LCCN or ISBN, and consequently, the link would be of little value for ILL. This arrangement would also make linked reference searching virtually impossible.

The Final Link Alternatives

Figure 5-4 lists the options that were selected.

FIGURE 5-4. OPTIONS REMAINING AFTER ELIMINATIONS

EXCHANGE METHOD

Direct Tape Delivery

Direct Batch Transmission

On-Line Access

DATA EXCHANGED

Requested Records

COMMUNICATION METHOD

Tape: Mail

Batch: Dial-up

On-Line: Leased Dedicated Lines

USER INTERFACE

Native Mode

Translation Mode

Six link alternatives can be constructed from these options:

(1) Direct tape delivery via mail of records requested in native mode

(2) Direct tape delivery via mail of records requested in translation mode

(3) Batch data transmission via dial-up line of records requested in native mode

(4) Batch data transmission via dial-up line of records requested in translation mode

(5) Online access in native mode via dedicated line

(6) Online access in translation mode via dedicated line

Three of these six alternatives were selected for an evaluation of their likely economic impact on the utilities and the benefits to libraries that each would provide. The results of this analysis are presented in Chapter 6.

It was decided that a detailed comparison should be made between the two online alternatives. An online link would give users the most convenient and immediate access to the resources and capabilities of other utilities. Further examination was needed to determine whether the native mode or translation mode version of an online link would be preferable. Two of the final three candidates, then, were the online access alternatives.

The tape delivery exchange method was preferred to the batch transmission alternatives because it was considered to be less complex technically. This apparent simplicity makes the tape delivery an attractive link mode if it can provide substantial benefits to the users. The native mode user interface was chosen for the tape delivery alternative because it would require less development effort than the translation option. Thus, the tape delivery/native mode alternative was the third candidate because it was thought to be the easiest to develop and the fastest way to establish a link.

In summary, the three alternatives selected for more detailed analysis are:

(1) Direct tape delivery via mail of records requested in native mode

(2) Online access in native mode via dedicated line

(3) Online access in translation mode via dedicated line.

These alternatives can be described as "user-initiated" because in each case the records are exchanged in response to a specific request by a user at another utility.

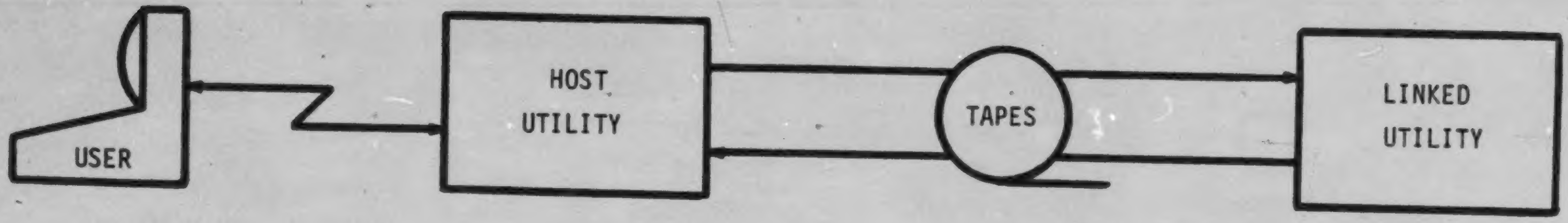
Figures 5-5, 5-6, and 5-7 describe for each alternative the steps that are involved from the time a user initiates a search until the user receives a response. Appendices 5A, 5B, and 5C are lists of the computer programs that would be required at each utility in order to implement each of the alternatives.

Direct Tape Delivery via Mail of Records Requested in Native Mode

In this alternative, if an item is not found on the host data base, the user enters a request and indicates that one or more of the other data bases is to be searched. The search commands are saved at the host utility and are periodically put onto a magnetic tape and mailed to the appropriate utility.

When a request tape is received, the utility runs the searches on its system, generates the responses in its format, and stores them on a tape. The program that processes the request tape will automatically provide the interactive responses that are necessary to retrieve all of the records that are found. (A limit would most likely be placed on the number of records that will be retrieved for each search.) The response tape is then mailed back to the requesting utility. The user's host utility loads the responses into an online file and makes any format adjustments necessary to permit the display of the records on the terminals of the host utility. The system notifies the user at sign-on that there is a response from a linked search waiting and sends the records to the terminal when the user requests to see them.

FIGURE 5-5. DIRECT TAPE DELIVERY VIA MAIL OF RECORDS REQUESTED IN NATIVE MODE



1. User searches host data base and then wishes to extend search to linked data base.
2. User enters request to search other data base(s).
3. System requests user to enter search statements in the format of the linked utility.
4. User enters search commands in response to prompts.
5. Requests from all users are stored in Request File, sorted by linked utility.
6. Periodically, requests for each destination are purged of duplicates, and a Request Tape is created for each linked utility.
7. Request Tapes are mailed.
8. Request Tapes are received and data is loaded into system.
9. Requests are processed.
10. Processing program provides continuation commands, but retrieves only a limited number of records for each search.
11. Responses are sorted by requesting utility and stored in Response File.
12. Periodically, a Response Tape is created for each requesting utility.
13. Response Tapes are mailed.
14. Response Tape is received and loaded into online Response File.
15. System notifies user at next sign-on that a response is waiting.
16. User asks to see waiting responses.
17. System sends responses to user's terminal.

The transmission time for this form of linking would be a week to 10 days if the U.S. Postal Service were used. This interval could be shortened, at added cost, by using a courier service. It is assumed that each utility would send a request tape to each linked utility once a week, and thus the actual delay could be 3 weeks or longer for a user who makes a request several days before the next request tape is scheduled to be mailed.

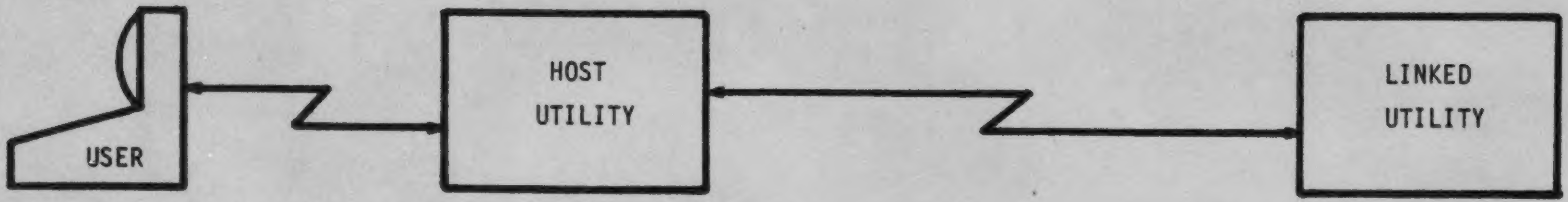
Online/Native Mode via Dedicated Line

In this scenario, after failing to find an item on the host data base, the user enters a command instructing the host utility computer system to provide access to one of the other data bases. The host computer then converts certain terminal commands from the user's terminal to the terminal commands of the linked utility. The terminal becomes in effect a part of the other network. The user is responsible for understanding the search commands and record formats of the linked utility.

The host computer receives the search request from the terminal and sends it to the linked utility, where it is read as though it were any other search from a participant in that network. The data base is searched, and the response is transmitted back across the line to the host utility's computer and then to the user's terminal. If further interaction is required to narrow the search or to page through a number of records, then this is accomplished in the same way.

When a record is displayed, the linked user has three options. (1) If the record is needed to perform copy cataloging, it may be printed

FIGURE 5-6. ONLINE ACCESS IN NATIVE MODE VIA DEDICATED LINE



1. User searches host data base and wishes to extend search to a linked data base.

2. User enters request to search other data base(s).

4. User enters search commands in the format of the other system.

11. User receives response and may print the record(s) for later rekeying. (Re-keying would not be necessary if computer software to translate responses were developed. See Chapter 6.)

3. System requests user to enter search statements in the format of the linked utility.

5. System converts terminal control characters and transmits request to the linked utility.

9. System processes response from other utility, converting terminal control characters, adjusting for screen size.

10. System transmits response to user's terminal in regular format of linked utility.

6. Requests are accepted, receipt is acknowledged, and requests are placed in queue.

7. Requests are processed online using existing search software.

8. Responses are transmitted to the requesting utility in the regular response format.

out on the user's printer. The information would then be entered in the host data base sometime later as a member-generated record. (Re-keying would not be necessary if computer software to translate responses were developed. This possibility is expanded in Chapter 6.) (2) The record may be reviewed visually for the needed information. For example, an interlibrary loan searcher might simply note one or more of the holding symbols. The record would not be printed. (3) The user could determine that the record is not useful and discard it immediately.

When the user is ready to terminate the link, a short command instructs the host computer to return the terminal to its normal status.

Online/Translation Mode via Dedicated Line

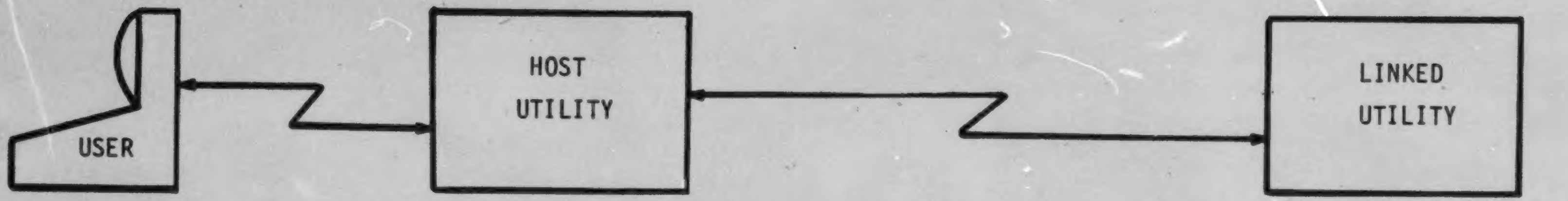
Where native mode requires the user to learn the search commands and argument formats of the other utility, the translation mode link would use a computer for at least part of the translation. In all other respects, this alternative is the same as the other online alternative.

There are at least two ways in which the translation of search requests might be achieved, and further study is needed before one can be selected as preferable. In one approach, suggested in an LC Network Development Office report on message text formats (1), the utilities would

---

(1) Philip L. Long, "Study of Message Text Formats: Bibliographic Search Queries", Network Planning Paper No. 5, Washington, D.C.: Library of Congress, 1979.

FIGURE 5-7. ONLINE ACCESS IN TRANSLATION MODE VIA DEDICATED LINE



1. User searches host data base and then wishes to extend search to linked data base.
2. User enters request to search linked data base(s).
3. System sends the user a "fill-in" screen, requesting additional information about author and title, if necessary.
4. User provides additional information where requested.
5. System converts request to a standard message format.
6. Request is transmitted to linked utility using NCLIS/NBS protocol or variant.
7. Requests are accepted and receipt is acknowledged.
8. System converts request from standard message format to regular search format, and requests are queued.
9. Search requests are processed online using existing search software.
10. Responses are converted to standard message format.
11. Responses are transmitted to requesting utility using NCLIS/NBS protocol or variant.
12. Responses are received and converted from standard message format to regular response format.
13. Responses are transmitted to the user's terminal.
14. User receives response and may modify online.

agree on standard formats for certain search arguments. A user who wishes to search another data base could receive a fill-in screen or some other form of prompt at the terminal asking for, at a minimum, (a) the data base to be searched, (b) the type of search (e.g., "find"), and (c) the standard form of the search argument. The host computer would then add the appropriate search command and suffixes to the argument and transmit it. The linked utility's computer would translate the argument from the standard format to the format used on that system.

Another approach to translation places more of the burden on the host utility's computer. This approach is depicted in Figure 5-7. Fully-automated translation could be performed for search arguments that are identical for at least two of the systems. The searcher would enter the request statement in the host utility's format, and the host computer would then replace the local search command with the appropriate command for the linked utility. LC Card Number and ISBN searches can be translated automatically now at all four utilities. Some other keys may be handled this way between at least two utilities--for example, title keyword searching between WLN and RLIN.

Sometimes, however, the user would need to supply more information so that the translation may be made. For example, the OCLC title search GON,WI,TH,W must be converted to a form such as FIN T GONE WIND to be an effective title search on RLIN. In this case, the user receives a fill-in-the-blanks request on the terminal screen and keys in the two or three most distinctive words in the title. By the same token, an RLIN user who does not succeed with FIN T GONE WIND and who wishes to search OCLC's data base is required to enter the first four words of the title. The RLIN computer would then automatically truncate the words to the 3,2,2,1 OCLC format and add the commas.

An elaboration of this concept allows users to select from a menu on their screens the type of search key they would like to use for the linked search. This enables users at one utility to take advantage of the unique search capabilities of the other systems. For example, an OCLC user may request to search RLIN or WLN by subject heading, a capability that is not currently available on OCLC. A fill-in screen would appear asking for the term or terms to be used in the search. This information is then converted automatically into the proper format for subject searching on RLIN or WLN, and the request is transmitted to the specified utility.

Much more work is required to determine the most practical way of translating responses to search queries. Standard formats must be developed for bibliographic records, perhaps based on the ANSI Z39.2-1979 standard, and also for error messages and statements about the result of a search.

CHAPTER 6

A COMPARATIVE EVALUATION OF THREE LINK ALTERNATIVES

Introduction

This chapter examines in greater depth the three link alternatives selected in Chapter 5: (a) tape delivery of requested records, (b) online access in native mode, and, (c) online access in translation mode. The benefits of each type of link are analyzed, and the costs to the utilities of developing and operating them are projected. The alternatives were evaluated in four areas:

- The benefits to libraries that use the resources made available through linking
- The computer resources required at each utility to support the anticipated increase in traffic
- The costs to each utility of the computer processing and data transmission associated with the link
- The staff resources needed for design, development, and implementation.

The comparison reveals that the online link with translation interface would be the most appealing to libraries, but that development and implementation would take the longest of any of the alternatives. The online/native mode alternative has the disadvantage of requiring that searchers be familiar with the commands and record formats of the linked utilities, but this type of link could be established within 3 years and could provide a foundation for development of an online link with automatic translation. The tape delivery of requested records is less desirable because delays of a week or more in the availability of records could substantially reduce the link's usefulness.

Portions of the analysis for this chapter were performed with the aid of a computer-based model developed by Battelle. The model, called BIBLINK, was designed to help the utilities and other interested parties in the ongoing process of assessing the implications of linking. BIBLINK was used to project the computer resources required, the cost of data processing, and the cost of data transmission for each of the three alternatives. Appendix 6A describes the model in greater detail.

Not all of the utilities provided sufficient information about their operations and plans to enable a reliable projection of their costs from linking. Battelle estimates have been used in some cases where data were not made available; in other cases, no projections were made at all, although the capability to do so exists in the model. The results of the cost analysis must therefore be viewed as a rough cut, subject to refinement if better input data are made available.

The Benefits to Libraries of Each Alternative

The analyses in Chapters 2, 3, and 4 indicate that linking has the potential to provide libraries with substantial benefits in the form of improved service and avoidance of processing costs. An important criterion used in comparing the link alternatives is the degree to which each alternative allows libraries to obtain these potential benefits. For example, the link alternatives will differ in the ease of use--the amount of special effort that is required of the people who will conduct linked searches. The ideal link alternative, from the library's point of view, would be one that makes searching linked data bases no more difficult than searching the host data base, while providing timely access to all of the unique resources not available on the host data base.

Timeliness of Access

All three of the link alternatives may be described as "user-initiated" links, because the bibliographic records are transferred only in response to a specific request. This type of link has the advantage of providing access to all of the resources on the linked data bases. As Chapter 5 noted, this is not true for link alternatives that involve replicating records (e.g., all new member-generated records) on each data base. Such a link would permit access to only the replicated records, and many other potentially useful records would remain inaccessible.

While each of the three alternatives provides full access to the resources of each data base, they should also be compared on the basis of the timeliness of that access. Here there is a clear difference. Both online alternatives offer "real-time" access; a linked search takes about the same amount of time at the terminal as a search of the host data base.

However, the tape delivery alternative would probably involve a delay of more than a week between requests and responses because the data are mailed. Assuming that request tapes are sent once a week, a searcher who submits a request early in the week may have to wait as long as 2 weeks--1 week while the requests are accumulating before being sent, and another week or so while tapes are in the mail.

This lag would inhibit the usefulness of the link. In particular, the delay would make linked searching less useful for verification, location, reference searching, and authority checking. Some ILL librarians might choose to use printed sources for verification and location before paying for a search of a linked data base that would provide results 1 or 2 weeks later. On the other hand, some people might be willing to wait this long for the results of a reference search--particularly a subject search--but again the delay would reduce the value of linking. Catalogers, who are used to waiting for a record to appear on the host data base, may accept the delay for bibliographic file searches, but the lag would discourage linked authority searching.

The delivery of the tapes could be speeded through the use of an express courier service (Federal Express, Emery Air Freight, etc.), but the delay would still be at least 3 days once the tapes were sent. Even if the utilities sent tapes once a day, the responses would not be received for 3 or 4 days. In any event, tape delivery of any sort is less desirable with respect to timeliness than the online alternatives because this type of link does not offer libraries the benefit of immediate access to the information of other utilities.

Ease of Use

The tape delivery alternative and one of the online alternatives employ the native mode interface, in which the searcher in effect does the translating instead of the computer. The searcher must be trained to use the command syntax and search strategies of each of the linked data bases and must also be able to interpret the response statements and record formats. If the library wishes to "capture" a record found on another data base, then the data must be re-keyed at the terminal and entered in the host data base as though a member-generated record were being created. This type of link requires more effort by the searcher than the translation interface, in which the link is virtually transparent. With the translation interface, the host utility's computer system automatically translates requests into a standard format or prompts the searcher for more information when direct translation is not possible. The responses from the linked utility are automatically converted from the standard format into the format of the searcher's host utility. If desired, a record resulting from a linked search can then be entered into the host data base without re-keying.

From the user's point of view, then, the translation interface is clearly preferred to native mode interaction because it does not require learning to search the linked data bases. Also, the user does not have to manually translate a record received in the linked utility's format and then re-key the data to add the record to the host data base. Preparing documentation and training users would not take as long as with a native mode arrangement.

Figure 6-1 summarizes the standing of the three alternatives on the two criteria discussed so far. The online/translation link scores well on both timeliness of access and ease of use. The online/native mode link provides timely access but is more difficult to use. The tape delivery alternative rates poorly on both criteria.

FIGURE 6-1. BENEFITS TO LIBRARIES FROM EACH LINK ALTERNATIVE

Link Alternative	Timeliness of Access	Ease of Use
Online access in translation mode	High	High
Online access in native mode	High	Low
Tape delivery by mail of records requested in native mode	Low	Low

Computer Resource Requirements for Each Alternative

A major concern of the utilities is the effect that the added load from linked users will have on the performance of their existing computer services to regular users. If a great number of linked searches must be processed at a given hour, to what extent will response time be degraded, creating a nuisance for all users? Does it appear that additional computer processing capacity is required to absorb the linked search load and preserve an acceptable response time? The data presented in this section can help the technical staffs of the utilities to answer these questions, using their own knowledge of the computer speeds, load distributions, and capacity available in each case.

Anticipated Search Loads from Linking

Figures 6-2, 6-3, and 6-4 show the estimated number of search requests that each utility might receive from linked users in an average day. These projections are based on data collected throughout the study and on assumptions about the willingness of library staffs to conduct linked searches for cataloging, authority checking, ILL verification and location, and reference searching. Appendix 6B explains the methods and assumptions used to derive the estimates.

The estimates also assume that a single item will produce only one search of the online data base. Actually, the average number of searches per item is greater than 1 because of typographical errors, multiple key searches, and other factors. However, no data were available to suggest an appropriate ratio, and so the lower bound was used here. As a result, the search loads--and the operations costs derived from them--are minimum estimates. The BIBLINK computer model may be used to produce projections based on searches-per-item ratios greater than 1.

Two sets of search load estimates are provided. One projects the traffic that would occur if a link were operational today. These figures represent the present demand for linked searching and are based on the findings presented in earlier chapters and on the current traffic at each utility. Since a link will require at least 2 years to implement, a second set of estimates is also provided. The 1983 estimates project the added

FIGURE 6-2. ESTIMATED NUMBER OF SEARCH REQUESTS RECEIVED EACH DAY FROM USERS OF OTHER UTILITIES

Link Alternative: Tape Delivery of Requested Records

1980	Shared Cataloging	Authority Checking	Interlibrary Loan	Reference Searching	Total
OCLC	1,080	-	15	15	1,110
RLIN	500	40	100	105	745
WLN	90	45	5	60	200
LC	245	110	-	-	355

1983	Shared Cataloging	Authority Checking	Interlibrary Loan	Reference Searching	Total
OCLC	1,240	-	15	15	1,270
RLIN	960	40	100	105	1,205
WLN	125	45	5	60	235
LC	390	110	-	-	500

See Appendix 6B for an explanation of how linked search loads were derived.

FIGURE 6-3. ESTIMATED NUMBER OF SEARCH REQUESTS RECEIVED EACH DAY FROM USERS OF OTHER UTILITIES

Link Alternative: Online Access in Native Mode

1980

	Shared Cataloging	Authority Checking	Interlibrary Loan	Reference Searching	Total
OCLC	1,080	-	85	855	2,020
RLIN	500	8,910	180	13,530	23,120
WLN	90	2,880	35	2,700	5,705
LC	245	-	-	-	245

1983

	Shared Cataloging	Authority Checking	Interlibrary Loan	Reference Searching	Total
OCLC	1,240	-	85	945	2,270
RLIN	960	11,140	245	15,490	27,835
WLN	125	3,825	55	3,060	7,065
LC	390	-	-	-	390

See Appendix 6B for an explanation of how linked search loads were derived.

FIGURE 6-4. ESTIMATED NUMBER OF SEARCH REQUESTS RECEIVED EACH DAY FROM USERS OF OTHER UTILITIES

Link Alternative: Online Access in Translation Mode

1980	Shared Cataloging	Authority Checking	Interlibrary Loan	Reference Searching	Total
OCLC	1,200	-	95	950	2,245
RLIN	570	9,900	205	15,825	26,500
WLN	100	3,200	35	3,000	6,335
LC	270	-	-	-	270

1983	Shared Cataloging	Authority Checking	Interlibrary Loan	Reference Searching	Total
OCLC	1,375	-	95	1,050	2,520
RLIN	1,065	12,375	270	17,210	30,920
WLN	140	4,250	60	3,400	7,850
LC	430	-	-	-	430

See Appendix 6B for an explanation of how linked search loads were derived.

search load that each utility could expect in what might be the first year of operation. These numbers are larger than 1980 estimates because they take into account the membership growth at each utility. However, it is not possible to foresee the new services that will be offered by the utilities 2 or 3 years from now, nor can changes in computer capacity and capabilities be predicted. The 1983 figures are not meant to be firmly associated with that year, but they are simply estimates of the magnitude of linked searching a few years from now.

The projected search loads are different for the three link alternatives. Since the native mode interface is more difficult for the searcher to use, it was assumed that the online/native mode link would be used less frequently than the online/translation link. The online/native mode search loads were calculated as 90 percent of those estimated for online/translation. The projections for the tape delivery alternative are much lower than those for either of the online alternatives, reflecting the assumption that the mail delay would substantially reduce the usefulness of the link.

A notable feature of these search load estimates is that RLIN could bear the major share of linked searching if an online link is implemented. The RLIN data base would in most cases be the first data base searched by OCLC libraries that do not find what they need on the OCLC data base. More than half of the linked searches on RLIN might be expected to be subject searches by OCLC libraries that are taking advantage of a feature not offered by their own utility. This projection is founded on speculation, though, since the demand for subject searching by OCLC members is not known. The RLIN figures also assume that RLG will develop an online authority file within the next year and that many OCLC users would access this file for authority checks.

Daily Computer Processing Requirements

Each utility provided estimates of the average computer time required by its system to process searches of bibliographic files and authority files. The estimates were based on the existing hardware and software configurations. (Appendix 6C lists the hardware and software at each utility at the time of this study). These data were used to calculate the total amount of computer time that would be needed to process the incremental search loads in Figures 6-2, 6-3, and 6-4.

The estimated computer-hours required each day at each utility are summarized in Figure 6-5. (A computer-hour is equivalent to the operation of one computer at full capacity for 1 hour.) These estimates combine the computer time needed to process incoming search requests from linked users and the time needed to prepare searches from the utility's own users to be sent out to other utilities. As with the search load estimates, these computer time projections are for the minimum level of linked searching that might be expected.

The computer time required for the tape delivery alternative is very small for RLIN, WLN, and LC. The estimate is much larger for OCLC, but it still appears to be small enough to be handled by the present OCLC facilities, especially if the request tapes are processed at night. This

FIGURE 6-5. COMPUTER-HOURS PER DAY REQUIRED TO PROCESS SEARCH LOADS DUE TO LINKING

1980			
	Tape Delivery	Online Native Mode	Online Translation Mode
OCLC	.54	2.62	2.91
RLIN	.02	1.13	1.25
WLN	.02	.69	.76
LC	.07	.05	.06

1983			
	Tape Delivery	Online Native Mode	Online Translation Mode
OCLC	.63	3.01	3.35
RLIN	.03	1.32	1.47
WLN	.03	.82	.91
LC	.09	.07	.08

Computer processing overhead is assumed to be 100% of applications processing time. Computer-hours per day = (Total computer-seconds for incoming searches + total computer-seconds for outgoing searches) x 2 for 100% overhead time ÷ 3600 seconds per hour.

Estimates of computer-seconds per average search were provided by the utilities. Additional time needed to translate requests and responses was estimated by Battelle.

pattern is repeated for the two online alternatives, reflecting the fact that OCLC's computer system takes more time to process a search request than those used by the other utilities. The Stanford University IBM 3033 computer used by RLIN processes searches with the least CPU time, and therefore the very large daily search loads expected for RLIN would take a relatively moderate amount of computer time.

The computer time estimates do not appear to be so high in any of the cases that they render an alternative impractical. The search load from the tape delivery alternative would be fairly light, and none of the utilities would have trouble accommodating this additional load. The additional computer time needed for the online alternatives is considerably greater, but only at OCLC would it exceed about 1 and a half computer-hours per day. OCLC would require more than 3 computer-hours to process the 1983 daily load from the online/translation link. The precise impact of 3 additional hours of computer processing must be determined by OCLC technical analysts, but as an example, if it were distributed evenly through a 10-hour day, this would mean that linked searching would require about 30 percent of the capacity of one OCLC computer in each of the 10 hours.

Incremental Costs of Computer Processing and Data Transmission

Figures 6-6 and 6-7 present estimates of the major operations costs of linking in each of the three ways. 1980 estimates have again been provided as the baseline from which the 1983 figures were derived. The cost of processing both incoming and outgoing requests was calculated by multiplying two factors: (1) computer time needed to process the estimated

FIGURE 6-6. INCREMENTAL ANNUAL COSTS DUE TO LINKING, 1980 ESTIMATES

Link Alternative:
Tape Delivery of Requested Records

	OCLC	RLIN	WLN	LC
Processing Incoming Search Requests	$138,900	$ 6,100	$ 4,600	$18,500
Processing Outgoing Search Requests	12,300	500	400	2,400
Data Transmission (mail)	1,100	1,100	1,100	1,300
Total	$152,300	$ 7,700	$ 6,100	$22,200

Link Alternative:
Online, Native Mode

	OCLC	RLIN	WLN	LC
Processing Incoming Search Requests	$323,000	$314,300	$158,800	$12,800
Processing Outgoing Search Requests	413,900	3,400	1,800	2,600
Data Transmission (leased line)	24,500	27,600	21,500	28,200
Total	$761,400	$345,300	$182,100	$43,600

Link Alternative:
Online, Translation Mode

	OCLC	RLIN	WLN	LC
Processing Incoming Search Requests	$358,900	$349,200	$176,400	$14,200
Processing Outgoing Search Requests	459,800	3,700	2,000	2,900
Data Transmission (leased line)	24,500	27,600	21,500	28,200
Total	$843,200	$380,500	$199,900	$45,300

Processing costs were derived by multiplying computer-seconds required for applications processing by $.50, the approximate cost to RLG and WLN of a computer-second.

Transmission costs for the online alternatives are based on the assumption that each pair of utilities will pay half of the cost of the line between them.

FIGURE 6-7. INCREMENTAL ANNUAL COSTS DUE TO LINKING, 1983 ESTIMATES

Link Alternative:
Tape Delivery of Requested Records

	OCLC	RLIN	WLN	LC
Processing Incoming Search Requests	$198,100	$11,100	$6,400	$32,600
Processing Outgoing Search Requests	23,300	500	800	3,600
Data Transmission (mail)	1,400	1,400	1,400	1,600
Total	$222,800	$13,000	$8,600	$37,800

Link Alternative:
On-Line, Native Mode

	OCLC	RLIN	WLN	LC
Processing Incoming Search Requests	$ 451,500	$460,200	$235,700	$25,400
Processing Outgoing Search Requests	609,200	5,200	3,000	3,800
Data Transmission (leased line)	30,600	34,500	26,900	35,300
Total	$1,091,300	$499,900	$265,600	$64,500

Link Alternative
On-Line, Translation Mode

	OCLC	RLIN	WLN	LC
Processing Incoming Search Requests	$ 501,500	$511,300	$261,900	$28,200
Processing Outgoing Search Requests	676,800	5,800	3,400	4,200
Data Transmission (leased line)	30,600	34,500	26,900	35,300
Total	$1,208,900	$551,600	$292,200	$67,700

Processing costs were derived by multiplying computer-seconds required for applications processing by $.625, a 25% increase over the $.50 used in the 1980 cost estimates.

Transmission costs for the online alternatives are based on the assumption that each pair of utilities will pay half of the cost of the line between them.

annual search load from linking, and (2) the estimated costs to utilities of a unit of computer time.

RLIN and WLN each use university computing facilities and are charged on the basis of central processing seconds used. OCLC owns most of its facilities, and the Library of Congress leases its computers, and so neither organization pays a variable usage charge like RLG and WLN. OCLC and LC could not provide the research team with information that would permit estimation of a unit cost of processing. Therefore, the $.50 used for RLIN and WLN was also used for OCLC and LC, on the assumption that if the actual values could be calculated, they would not differ greatly from this figure. The operations costs for OCLC and LC in Figures 6-6 and 6-7 should be interpreted with this qualification in mind.

The data transmission cost for the tape delivery alternative is simply the cost of mailing two tapes each week to each of the linked utilities--one tape containing search requests and the other containing the responses to the previous week's requests. Requests and responses could be put on a single tape in many cases, and so two tapes a week is a liberal estimate. The transmission cost for the online alternatives is the shared cost of a dedicated leased line between each pair of utilities. It has been assumed that each pair of utilities would evenly divide the monthly charge for the line connecting them.

The data processing and transmission costs are not the only recurring costs associated with linking. The expenses for additional operations staff, maintenance, and overhead must be considered. Some account should also be taken of the revenue lost, if any, when operations time is reallocated to linking from other revenue-generating applications. These costs were not included in the analysis in this chapter because none of the utilities provided the project team with information on which meaningful estimates could be based. The BIBLINK computer model was designed to accommodate input data in these areas, though, and this tool may be used to expand the cost analysis as the appropriate data become available.

Tape delivery is by far the least expensive of the three link alternatives, and this is due to the low projected search loads and the minor cost of postage. As explained earlier, though, this low cost is accompanied by reduced benefits to the libraries using the link. If a courier service were used to cut the delay to 3 or 4 days, the cost would be about $2,500 more a year for each of the utilities in 1983.

The online/translation alternative is the most expensive to operate. It would cost WLN about $292,000 to process all of the search requests projected for 1983. RLIN would receive a very large number of requests from other utilities, and the cost of computer processing would be high if the full demand for linked searching were met. In all, linked searching could cost RLIN more than $550,000 in 1983.

The operations costs projected for OCLC and LC are less reliable because no information was provided regarding the unit cost of computer time at these organizations. At a rate of about $.63 per computer-second, linking would cost OCLC over $1.2 million in 1983. This figure is much greater than the others because of the comparatively large amount of computer time needed to process a search of the OCLC data base. Using the same unit rate, $67,700 is projected as LC's cost of linking in 1983.

The cost to all four utilities is lower for the native mode alternative than for the translation link because the search load for native mode was assumed to be only 90 percent of that estimated for the translation alternative.

It must be re-emphasized that these cost projections are based on the minimum number of searches that could result from the link traffic estimated in Figures 6-2, 6-3, and 6-4, where one search per item was assumed. The estimates can be revised upward in direct relation to the searches-per-item ratio. For example, if new data indicate that an average of two searches is conducted per item, then the costs in Figures 6-6 and 6-7 should all be doubled.

Development Effort

Staff Time Requirements

Figure 6-8 summarizes the estimated development effort in person-years that would be required to implement each of the three link alternatives. These estimates were derived by summing the staff time estimates for each task defined in the more detailed implementation plans presented in Appendices 6D, 6E, and 6F.

The calendar year estimates in Figure 6-8 are conjectural. The number of staff assigned by each utility will greatly affect the time required to implement the link. The projections made here--1.5 years for

FIGURE 6-8. ESTIMATED DEVELOPMENT EFFORT FOR EACH LINK ALTERNATIVE

		Planning and Defining Standards			Designing, Programming, and Testing			Total		
		Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years
TAPE DELIVERY	Applications Software	.25	1	.5	2	8	1	2.25	9	1.5

		Planning and Defining Standards			Designing, Programming, and Testing			Total		
		Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years
ONLINE ACCESS IN NATIVE MODE	Applications Software	.5	2	.75	2.5	10	1.5	3	12	2.25
	Communications Software	1	4	1	2-4	8-16	1.5-2	3-5	12-20	2.5-3
	Total	1.5	6	1	4.5-6.5	18-26	1.5-2	6-8	24-32	2.5-3

		Planning and Defining Standards			Designing, Programming, and Testing			Total		
		Person Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years
ONLINE ACCESS IN TRANSLATION MODE	Applications Software	1	4	1	4	16	2	5	20	3
	Communications Software	2	8	1	2-4	8-16	2-3	4-6	16-24	3-4
	Total	3	12	1	6-8	24-32	2-3	9-11	36-44	3-4

Person-Year = The equivalent of one person working full-time for one year.

Calendar year estimates are based on the assumption that each utility assigns 3 full-time equivalent staff to the project.

Development of communications interface is assumed to occur independently at each utility.

tape delivery, 2.5 to 3 years for online/native, and 3 to 4 years for online/translation--are based on the assumption that 3 full-time staff members are assigned by each utility.

The tape delivery alternative requires the least development effort, about 9 person-years altogether. No communications interface programs are required, and the existing search programs may be adjusted to permit the processing of request tapes. The major development effort would be the creation of a computer program at each utility to permit responses from a linked utility to be displayed on the host utility's terminals (e.g., convert terminal control characters, convert line-by-line format to screen format and vice versa).

Online/native mode uses existing applications software to service user search requests but the new response display program needed for the tape delivery alternative is also required. For applications software, online/native mode will require an estimated total of 12 person-years of effort, compared to the 9 person-years for the tape delivery alternative. The major difference in the two development efforts is that the online link requires communications interface software. It is assumed that the utilities will develop a subset of the NCLIS/NBS protocol for this purpose, and this is likely to require 3 to 5 person-years per utility, or 12 to 20 person-years altogether.

Online/translation mode requires the largest software development effort because this link employs standard request and response formats. Additional software is required to enable the computer to translate requests and responses automatically. The applications software development effort is estimated to be nearly double that of online/native mode--20 person-years versus 12 person-years. A more complete version of the NCLIS/NBS protocol would be desirable for the online/translation link, and the effort required to develop and implement this protocol is 4 to 6 person-years per utility, or 16 to 24 person-years altogether.

Earlier in the chapter, it was noted that an online/native mode link would be more difficult to use than a link providing automatic translation. With the native mode interface, the user must enter search commands in the format of the linked utility, and if a record found elsewhere is to be captured for the host data base, then the user must somehow copy the record from the terminal, translate the format manually, and re-key the data into the host data base as though it were a member-generated record. Part of this disadvantage could be avoided by developing a "hybrid" interface in which search requests are still entered in native mode but responses from the linked utility are translated by the host utility's computer. The computer programs and standard formats needed for this capability are the same as those that would be used in the full translation interface. The hybrid alternative would require more development effort than the native mode interface, adding perhaps 2 person-years to the 12 shown in Figure 6-8 for developing applications software for the online/native mode link. The 14 person-years is still much less effort than the 20 person-years required to develop the translation interface for both requests and responses.

It was assumed in the preceding analysis that each utility would develop communications interface software independently. This assumption was made because three different communications processors are now being used among the four organizations, and each computer would have its own special requirements if used as a link interface. The duplication of effort at the utilities might be avoided, however, if the same type of communications processor were to be used at each site. The communications interface software for the link, and part of the applications software, could be developed only once on a jointly selected minicomputer. A computer with a standard IBM or IBM-compatible host interface could be chosen so that three of the four utilities would not need to develop a new host interface.

The work could be done by a computer contractor, by a team of technical experts from each utility, or more likely, by a combination of utility and contractor representatives. Each utility would purchase a computer of the specified type and install the link software on it. Later participants in the linked network--institutions not considered in this study--would find it relatively easy to implement the communications interface portion of a link because they could simply buy the small computer and obtain a copy of the communications software.

Figure 6-9 provides estimates of the development effort required for this "common processor" approach. For the online/native mode link, the staff time would be 14 or 15 person-years, compared to 24 to 32 person-years if the interfaces are developed separately (see Figure 6-8). The online/translation link would require 24 or 25 person-years if a common processor is used, far less than the estimate of 36 to 44 person-years for independent development. Some of the effort avoided would be replaced by the staff time needed to select a mutually acceptable computer, and the

FIGURE 6-9. ESTIMATED DEVELOPMENT EFFORT IF COMMON COMMUNICATIONS PROCESSOR IS USED

			Planning and Defining Standards			Designing, Programming, and Testing			Total		
			Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years
ONLINE ACCESS IN NATIVE MODE	**Applications**	Utility	.5	2	.75	.75	3	1.5	1.25	5	2.25
		Contractor	-	-		-	2			2	
	Communications	Utility	.5	2	1	-	-	1.5-2	.5	2	2.5-3
		Contractor	-	2		-	3-4		-	5-6	
	Total	Utility	1	4	1	.75	3	1.5-2	1.75	7	2.5-3
		Contractor	-	2		-	5-6		-	7-8	

			Planning and Defining Standards			Designing, Programming, and Testing			Total		
			Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years	Person-Years per Utility	Total Person-Years	Calendar Years
ONLINE ACCESS IN TRANSLATION MODE	**Applications**	Utility	1	4	1	2	8	2	3	12	3
		Contractor	-	-		-	3		-	3	
	Communications	Utility	.5	2	1	-	-	2-3	.5	2	3-4
		Contractor	-	3		-	4-5		-	7-8	
	Total	Utility	1.5	6	1	2	8	2-3	3.5	14	3-4
		Contractor	-	3		-	7-8		-	10-11	

Person-Year = The equivalent of one person working full-time for one year.

Calendar year estimates are based on the assumption that each utility assigns 3 full-time equivalent staff to the project.

investment in the hardware would be an additional cost. However, if the utilities are willing to cooperate in this way, the common processor approach holds considerable promise for reducing development costs.

Staff Cost Estimates

Figure 6-10 shows estimates of the systems staff costs of developing and implementing the three linking alternatives. It is assumed that a communications interface will be developed separately by each utility. An average salary-plus-benefits cost of $30,000 is used to derive the total cost figures. This staff cost estimate is based on a project team of one systems programmer at $35,000/year, one systems designer at $30,000/year and one systems analyst at $25,000/year. The actual salary costs may vary widely from utility to utility because of regional differences in compensation.

The tape delivery alternative could be implemented for a total staff cost of about $270,000, an average of $67,500 for the 1.5-year effort at each utility. The cost for the online alternatives has been estimated in a range because of the difficulty of anticipating the nature of the protocol that would be implemented. The native mode link would require $180,000 to $240,000 in staff costs per utility, an overall cost of around $800,000. For the online/translation link, the staff costs would probably exceed $1,000,000 for the project as a whole, an average of $270,000 to $330,000 per utility.

FIGURE 6-10. PROJECTED SALARY EXPENSE FOR DEVELOPMENT OF EACH LINK ALTERNATIVE

	Development Effort in Person-Years		Salary Expense*	
Link Alternative	Each Utility	Total	Each Utility	Total
TAPE DELIVERY	2.25	9	$ 67,500	$ 270,000
ONLINE/ NATIVE MODE	6-8	24-32	$180,000 to $240,000	$ 720,000 to $ 960,000
ONLINE/ TRANSLATION MODE	9-11	36-44	$270,000 to $330,000	$1,080,000 to $1,320,000

See estimated development efforts in Figure 6-9.

* Development effort costed at $30,000 per person-year. This is an average of the annual salary expenses and benefits for one Systems Analyst at $25,000, one Systems Designer at $30,000, and one Systems Programmer at $35,000.

The non-salary costs of developing a link have not been estimated. Accurate projection of computer time, hardware, travel, and overhead costs should be based on a more detailed analysis than that undertaken for this study.

Summary of the Evaluation

Figure 6-11 summarizes the comparison of the link alternatives discussed in this chapter. The tape delivery alternative would require the least development effort, would consume minimal computer processing time, and would be the least expensive to operate. However, these advantages are gained at the expense of usefulness to libraries. The delay of a week or more between request and response would probably limit linked searching severely at many libraries. Both of the online alternatives are more beneficial because they would permit immediate access to the resources of the other data bases.

The costs to the utilities of establishing and operating an online link do not appear to be prohibitive. The translation link is the preferable alternative from the standpoint of user benefits, because a link of this type would be very easy to use, requiring only occasional entry of additional information to aid the translation of the commands that were used to search the host data base. User training would be much simpler than that required for native mode interaction. The principal drawback of the online/translation link is the development time. At the assumed staffing level of three full-time equivalent professionals per utility, linking would not become a reality until 1984 or 1985.

FIGURE 6-11. SUMMARY OF COMPARATIVE EVALUATION OF THREE LINK ALTERNATIVES

Criterion	Tape Delivery of Records Requested in Native Mode	Online Access in Native Mode	Online Access in Translation Mode
Timeliness of Access	Poor	Good	Good
Ease of Use	Poor	Fair	Good
Computer Time Required to Process Linked Searches	Easily absorbed by present facilities	Moderate burden	Moderate burden
Operations Costs	Low	High	High
Development and Implementation Effort	• 9 person-years • 1.5 calendar years • $67,500 per utility	• 24-32 person years • 2.5-3 calendar years • $200,000 per utility	• 36-44 person-years • 3-4 calendar years • $300,000 per utility
Summary Evaluation	Least desirable alternative. Low cost and low development effort offset by limited usefulness to libraries.	Native mode interface is burdensome to user, but should be considered as an interim arrangement leading to a translation link.	The preferred alternative, but will take more than 3 years to establish.

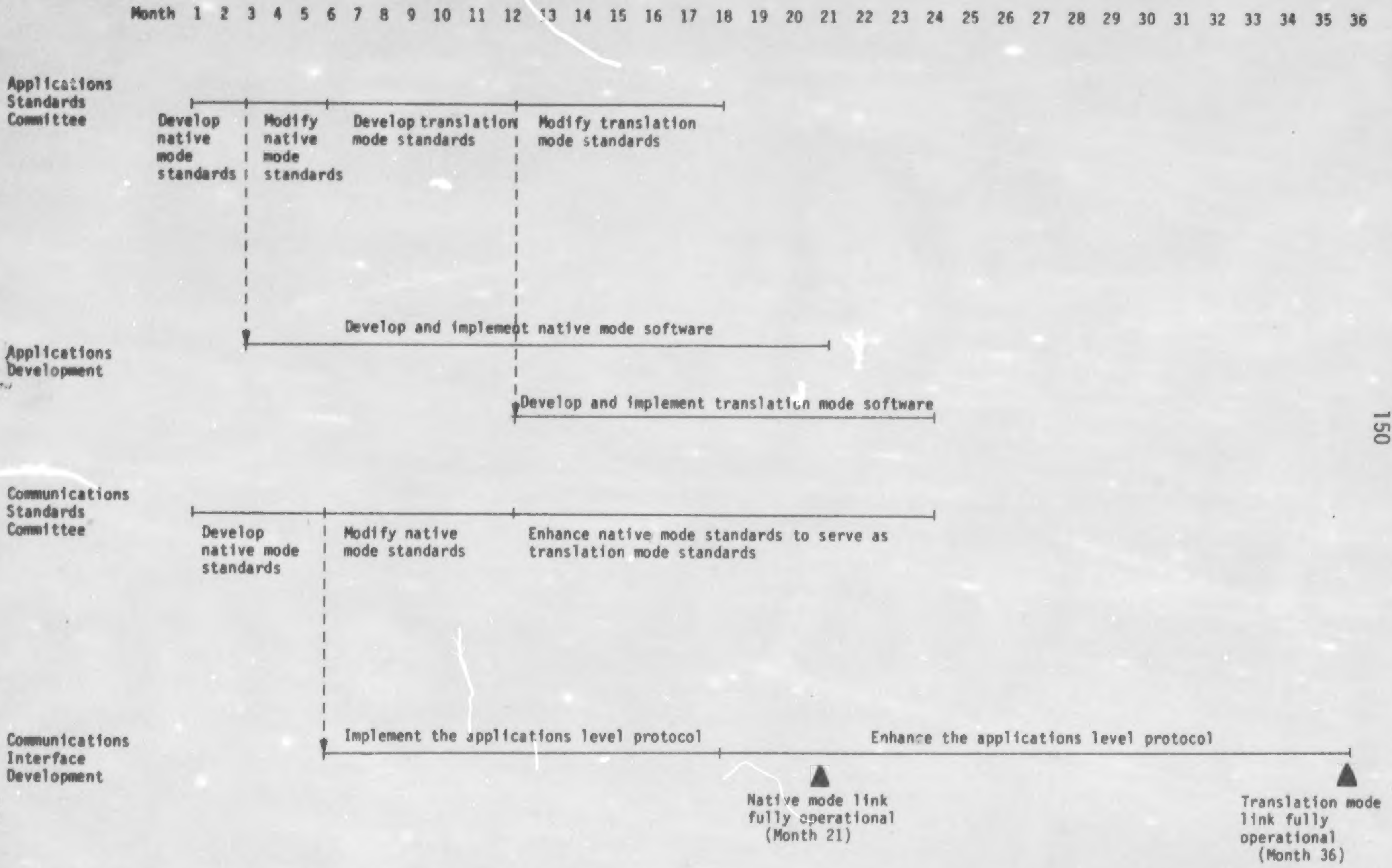
Implementation Schedule

The evaluation indicates that an online/translation link should be the goal of a program to link the utilities. Substantial effort is required to implement this type of link, though, in terms of both costs and calendar time, and the online/native mode alternative could be considered as an interim approach. A link could be established as much as 18 months sooner if the native mode interface is used, and development of the translation interface could begin even before the native mode link becomes operational. However, much of the native mode software would not be applicable to a translation mode link, and this wasted effort would be one significant price of implementing an interim link.

Figure 6-12 presents a simplified schedule for implementing a linked network. Appendices 6D, 6E, and 6F are implementation plans for each link alternative. Figure 6-12 was derived from these task breakdowns. This schedule should be viewed as just one alternative. Another possibility, not developed here, would be to phase in the translation mode link from the start, perhaps first supporting bibliographic searching and later supporting ILL as well.

The implementation plans assume that two standards committees will be formed. An Applications Standards Committee would define the user interfaces. For the native mode interface, the committee would determine how terminal control commands must be translated to permit a response from one utility to be displayed on another utility's terminals. The committee

FIGURE 6-12. IMPLEMENTATION SCHEDULE SHOWING OVERLAP OF DEVELOPMENT ACTIVITIES



See Appendix 6D for detailed implementation plans.

165

would then define the manner of entering requests for the translation link, including the fill-in screen used to prompt the user for more information, and would also determine the standard format into which requests and responses are to be translated before they are sent to a linked utility. A Communications Standards Committee would determine the exact nature of the interface protocols to be used for both the native mode and translation link types. Groups of utility representatives have already investigated both applications and communications standards for the exchange of authority data, and the efforts of these committees might be drawn on in linking the bibliographic files as well.

These two committees may perform many of their tasks concurrently with the implementation efforts. In Figure 6-12, the completion of the initial native mode application standards is followed by the development of the applications programs _and_ further work by the Applications Standards Committee, first on refined standards for native mode, and then on standards for translation mode. The same is true for the communications standards. After the initial interface standards have been developed, a subset of the applications protocol is implemented, while the Communications Standards Committee goes ahead with work that will lead to implementation of a full protocol for the translation link.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

Summary of Major Conclusions

The Benefits of Linking for Shared Cataloging

Hit Rates for Current Title Cataloging. A test search conducted in December 1979 and January 1980 suggested that linking the OCLC, RLIN, WLN, and Library of Congress data bases can be of significant value to libraries in cataloging current titles, providing access to many bibliographic records that would not have been available otherwise. Eight weeks after the titles were first searched on each data base, the hit rates were: 93 percent on OCLC, 87 percent on WLN and RLIN, and 86 percent on LC. The current title hit rate for all four data bases combined was 96 percent. Thus, catalogers using RLIN, WLN, or LC could have added about 10 percentage points to their current title hit rates if the four data bases had been linked at the time of the study. Most of the gain in each case would have come from the link with OCLC's data base. Catalogers in OCLC libraries could have gained about 3 percentage points on their current title hit rate if given access to the resources in the three other data bases.

The advantage from linking was greatest for research libraries and some other academic libraries, many of which could have added 15 to 20 percentage points to their current title hit rates. The incremental hit rate was less for public libraries, but many would have experienced an increase of 5 to 10 percentage points at the time of the study.

Avoidance of Cataloging Costs. Many libraries can derive substantial economic benefit if the utilities are linked. The added availability of bibliographic records means that libraries may avoid some of the effort, and thus some of the cost, of original cataloging. Calculations in Figure 2-16 indicate that if all four data bases had been linked in early 1980, RLIN participants could have avoided over $.80 for every current title cataloged. WLN users could have avoided nearly $.70 per current title and OCLC members about $.25 per current title. Altogether, linking could have resulted in the avoidance of nearly $3.4 million in cataloging costs in 1980.

Retrospective Conversion and Authority Work. Data collected and discussions with library professionals throughout the project suggest that linking can also aid shared cataloging in the areas of retrospective conversion and authority work, although neither topic was covered in detail by the study. As with current title cataloging, the benefit would take the form of an avoidance of some of the cost of doing original work by copying or otherwise using a record available in a linked data base. It appears that some libraries might drastically reduce their cost of authority work through linking, although the magnitude of the reduction cannot be estimated without research similar to that conducted for current titles in this study.

The Benefits of Linking for Interlibrary Loan

Effects on Lending. Linking the utilities would have little direct effect on lending operations. ILL librarians generally assume that lending requests have been verified by the sender, and a utility is not normally used in processing lending requests. However, linking may have an indirect effect on lending by making a library's holdings known to libraries that use other utilities. In some cases, this additional exposure could create an increase in the number of lending requests received, with a corresponding effect on ILL staffing and costs.

Verification of Monograph Borrowing Requests. A test search of monograph borrowing requests in February 1980 revealed that many libraries could verify more titles online if the utilities were linked. WLN and RLIN users could have performed online verification for an additional 35 percent of their requests, with most of this increase attributable to the resources at OCLC. OCLC users could have added only about 3 percentage points to their "verification hit rate", a modest gain. However, OCLC libraries might receive a different type of verification benefit if the utilites are linked. Some erroneous titles that cannot be verified on OCLC because of the relatively rigid OCLC command format might be verified on RLIN or WLN, both of which allow title word searching.

Location of Requested Items. The results of the test search also indicate that at least one potential lender can be found online for many more items if the utilites are linked. In early 1980, linking all four data bases would have enabled libraries to use a terminal to locate about 79 percent of their requested items. This represents an increase in the "holdings hit rate" of 46 percentage points for WLN and RLIN libraries, but only 2 percentage points for OCLC libraries. The increment is small for OCLC libraries because bibliographic records with holdings symbols attached are available on OCLC for about three-fourths of the titles requested, and linking improves this initial hit rate only slightly.

Avoidance of ILL Costs. Linking for ILL verification and location would enable many libraries to save time by verifying and locating online some of the titles for which records were not available in the host utility's data base. This time savings may result in an avoidance of some of the staff cost for ILL processing. The economic benefit from linking for ILL appears to be generally insignificant for the individual library, however, because of the small number of borrowing requests received by many libraries. Calculations in Figures 3-9 and 3-10 show that RLIN and WLN libraries might avoid more than $1.00 for every request processed if the four data bases were linked. OCLC libraries might avoid less than $.10 per request, though, because of the high initial hit rate on OCLC and the small increase in hit rate that can be obtained by linking. In all, linking could enable the library community to avoid about $75,000 a year in ILL processing costs.

The Benefits of Linking for Reference Searching

Replacing Local Catalogs With Online Union Catalogs. It is not presently feasible for OCLC to support the full volume of catalog use for all of its member libraries, whether or not linking is used as a means for extending the coverage of the online catalog. If OCLC's union catalog were to replace local catalogs, roughly 92 additional computer-hours would be required each day to process the online reference searches. This would require OCLC to add five computers to its facilities, an impossible burden under current circumstances. RLIN and WLN, both with far fewer participating libraries than OCLC, would not find the requirements overwhelming. Each would need about 1.5 additional computer-hours a day to provide online access in place of the local catalogs.

These estimates do not include the catalog use requirements of special libraries affiliated with any of the utilities.

Additional Successful Reference Searches. Linking the utilities could enable library patrons and staff to conduct over 2 million more successful reference searches per year than is possible when each library has access to only one online union catalog. The value of linked reference searching would be greatest for research libraries, with each library conducting perhaps 30 to 40 successful searches a day on a second data base.

These projections probably underestimate the value of linked reference searching. In this study, only searches that fail at the local catalog--searches that do not provide any useful citations--were counted as candidates for linked online searching. No data were available to permit projections of the full demand for online reference searching, including, for example, online searches for the purpose of expanding the results of a successful local catalog search. The true benefit from linked reference searching, measured in added information available to patrons and staff, would probably total several million successful searches a year.

The Method of Linking

Network Configuration. Each of the four organizations examined in this study--OCLC RLG/RLIN, WLN, and LC--has data, services or capabilities that could be useful to libraries affiliated with another institution. To maximize the benefit from linking, all four of the data bases should be linked. Users of each data base should have access to the unique features available from each of the others.

Type of Link. The most effective type of link between the four data bases would allow each utility user to request specific searches of other data bases. In other words, the ideal link would be "user-initiated". This arrangement is preferred to the replication of bibliographic records on multiple data bases. Replication alternatives are either inefficient because of the large number of records that need to be duplicated, or they do not permit complete sharing of bibliographic data because it is difficult for each utility to determine just which records will be useful to libraries using other utilities.

The Preferred Link Alternative. The preferred form of link is online access with automatic translation of request statements and responses using a dedicated leased line as the telecommunications channel. This alternative was chosen over others that were analyzed because it allows the user to access the resources of other utilities immediately, using the host utility's request statements, and then displays the responses in the host utility's record format. The link is nearly transparent, although the user may be required to enter some additional information at the terminal to aid the translation. Other alternatives that were considered feasible did not provide this degree of convenience or immediacy of access.

The Impact on the Utilities

Development of the Online Link. The online/translation mode link could be implemented in 3 to 4 years if the NCLIS/NBS protocol can be used and if each utility assigns three full-time equivalent staff members to the project. The staff salary expense for developing the link would be approximately $300,000 at each utility.

Computer Resources Required. When an online link becomes operational, RLIN would receive the greatest volume of linked searches by far--about 30,000 a day at a minimum. More than half of these incoming searches would be reference searches conducted by OCLC members taking advantage of RLIN's subject searching capability. With the present computer configuration, this added load would require less than 2 computer-hours per day in addition to the time required to process the regular load from RLIN participants. WLN would receive a minimum of about 7,000 linked search requests per day and would need less than 1 additional computer-hour to handle this load. The linked search load at OCLC is projected to be between 2,000 and 3,000 searches per day, requiring 3 to 3.5 additional computer-hours a day with the present configuration. The Library of Congress might receive 300 to 400 searches a day, assuming that only catalogers would access the LC bibliographic data. This added processing load would require about 0.1 computer-hours a day at LC.

The technical staff at each utility may use these projections to estimate the computer capacity required and the impact of linking on the response time for regular users. Also, the BIBLINK computer model may be used to produce projections based on searches-per-item ratios greater than 1.

Incremental Operations Costs. In 1983, probably the earliest point at which an online/translation mode link could be inaugurated, the additional computer processing and data transmission associated with linking would cost RLG about $550,000 for the year. At WLN, the annual increment in operations cost would be about $292,000. These projections are based on the assumption that RLG and WLN will pay for computer processing service by the computer-second, as they do now.

OCLC and the Library of Congress own or lease their own computer facilities and thus do not have an out-of-pocket expense for computer time. Neither organization provided information on its allocated cost per unit of time, and so Battelle used the rate reported by RLG and WLN to calculate the operations costs for OCLC and LC as well. The incremental cost to OCLC was projected at about $1,209,000 for 1983. LC would incur linking costs of about $67,700. More reliable estimates must be made when better data are available.

Recommendations

Continue Analysis and Planning Using the BIBLINK Computer Model

BIBLINK, the computer-based simulation model developed by Battelle, should be used by the utilities and by other interested parties to expand the analyses presented in this report and to explore issues that were not addressed here. The model is designed to aid the utilities in the areas of facilities planning, budgeting, and pricing policy. BIBLINK users may incorporate their own assumptions about the future volume of linked searching, and they may also supply some of the technical and financial data that were not available to Battelle at the time of the study. The model should be maintained online at a central facility and should be made available to remote users via dial-up terminal. Full security must be provided so that each user may protect sensitive data.

Develop an Online Link with Automatic Translation of Requests and Responses

Work should begin immediately to connect the host computers used by OCLC, RLG, WLN, and the Library of Congress. The goal of the linking project should be to implement an online "translation mode" link of bibliographic and authority files by the end of 1984. Development of the link should be undertaken jointly by the three utilities and LC. The work already performed by RLG, WLN, and LC on the Linked Authority System Project can be viewed as the initial steps toward this goal, and the agreement to cooperate on LASP might serve as a framework for further efforts involving OCLC.

Appoint Committees to Establish Standards

One of the earliest steps in the development of an online link is to form two committees. These bodies should include technical representatives from each utility and the Library of Congress. An Applications Standards Committee should determine the requirements for the user interface, particularly the standard formats to be used in the translation of requests and responses. This group should also devise a method of accounting and billing for linked searching. The CLR Task Force on a National Authority File Service might, with an expanded charge, serve this purpose.

A Communications Standards Committee should investigate the suitability of the NCLIS/NBS protocol for this application and determine the types of communications hardware and software required. This committee should also examine the feasibility of using a common communications processor at all four sites, thereby eliminating the need to develop complicated communications interfaces separately at each utility. Development time and cost should be reduced significantly if the utilities can cooperate in this manner. Another early task of the Communications Standards Committee should be to determine whether conditions favor the use of a value-added network instead of dedicated leased lines as the means of connecting the host computers.

BIBLIOGRAPHY

Abrams, Marshall D. and Cotton, Ira W., Introduction to Computer Networks: A Tutorial, Chevy Chase, MD: Computer Network Associates, 1976.

Allison, Anne Marie and Allan, Ann, OCLC: A National Library Network, Short Hills, New Jersey: Enslow Publishers, 1979.

Anderson, LeMoyne W., ed. Networks and the University Library, Chicago: Association of College and Research Libraries, 1974.

Aronofsky, Julius S. and Korfhage, Robert R., "Telecommunications in Library Networks: A Five-Year Projection", Journal of Library Automation, 10:1 (March 1977), pp. 5-27.

Arthur D. Little Inc., A Comparative Evaluation of Alternative Systems for the Provision of Effective Access to Periodical Literature, Washington, DC: National Commission on Libraries and Information Science, October 1979.

Avram, Henriette D., MARC: Its History and Implications, Washington, DC: Library of Congress, 1975.

Avram, Henriette D., "The National and International Network and Urban Main Libraries", in Paxton Price, ed., Future of the Main Urban Library, Las Cruces, New Mexico: Urban Libraries Council, 1978.

Avram, Henriette D., "Toward a Nationwide Library Network", Journal of Library Automation, 11:4 (December 1978), pp. 285-298.

Avram, Henriette and Hartmann, David, "Objectives and Accomplishments of the Network Technical Architecture Group", Program, 13:1 (January 1979), pp. 1-13.

Avram, Henriette and Maruyama, Lenore S. ed., Toward a National Library and Information Service Network: The Library Bibliographic Component, Preliminary edition, Washington, DC: Library of Congress, June 1977.

Bahr, Alice Harrison, Automated Library Circulation Systems 1979-80, White Plains, New York: Knowledge Industry Publications, 1979.

Bierman, John K., "Automated Alternatives to Card Catalogs", Journal of Library Automation, 8:4 (December 1975), pp. 277-298.

Bookstein, Abraham, "Congestion at Card and Book Catalogs--A Queuing-Theory Approach", Library Quarterly, 42:3 (July 1972), pp. 316-328.

Braunstein, Yale M., "Costs and Benefits of Library Information: The User Point of View", Library Trends, 28:1 (Summer 1979), pp. 79-87.

BIBLIOGRAPHY (Continued)

Brown, M. K. and McHugh, A. L., *Survey of Costs in Technical Processing and Interlibrary Loan*, Boulder, Colorado: Western Interstate Library Coordinating Organization (WILCO), December 1976.

Buchinski, Edwin J., *Initial Considerations for a Nationwide Data Base* (Network Planning Paper No. 3), Washington, DC: Library of Congress, 1978.

Buckland, Lawrence F. and Basinski, William L., *The Role of the Library of Congress in the Evolving National Network*, Washington, DC: Library of Congress, 1978.

Butler, Brett, *A Nationwide Location Data Base and Service* (Network Planning Paper No. 1), Washington, DC: Library of Congress, 1978.

Butler, Brett, West, Martha W., and Aveney, Brian, *Library and Patron Response to the COM Catalog: Use and Evaluation*, Los Altos, California: Information Access Corporation, 1979.

Cornew, Ronald W. and Morse, Philip M., "Distributive Computer Networking: Making It Work on a Regional Basis", *Science*, 189:4202 (August 15, 1975), pp. 523-531.

Cossar, Bruce, "Interlibrary Loan Costs", *RQ*, (Spring 1973), pp. 243-246.

Costello, Joseph G. and Cornog, Martha, *Library of Congress Data Collection Survey: Final Report*, Philadelphia, Pennsylvania: Auerbach Associates, 1978.

Costello, Joseph G. and Cornog, Martha, *Library of Congress Data Collection Survey: Non-Respondent Characterization Report*, Philadelphia, Pennsylvania: Calculon, 1979.

Cotton, Ira W., *Computer Science and Technology: Computer Network Interconnection: Problems and Prospects*, Washington, DC: National Bureau of Standards, 1977.

Daly, Richard R., *The Future Crisis in the Card Catalog: How Libraries Are Preparing for It*, Media, Pennsylvania: Information Associates, 1979.

Data Use and Access Laboratories, *Report on the Conference on Cataloging and Information Services for Machine-Readable Data Files*, Arlington, Virginia, 1978.

Dataflow Systems, *A Glossary for Library Networking* (Network Planning Paper No. 2), Washington, DC: Library of Congress, 1978.

BIBLIOGRAPHY (Continued)

De Gennaro, Richard, "Research Libraries Enter the Information Age", *Library Journal*, 104:20 (November 15, 1979), pp. 2405-2410.

Debenham, W. Stuart, Rastogi, Kunji, and Schieber, Philip, *OCLC Public Use Terminals: Report of Survey of Users of OCLC Public Use Terminals 1974-1975*, Columbus, Ohio: Ohio College Library Center, June 1976.

Dodd, Sue A., "Building an Online Bibliographic/MARC Resource Data Base for Machine-Readable Data Files", *Journal of Library Automation*, 12:1 (March 1979), pp. 6-21.

Dranov, Paula, *Automated Library Circulation Systems 1977-78*, White Plains, New York: Knowledge Industry Publications, 1977.

Eckard, Helen M., *Survey of Public Libraries*, *LIBGIS I*, Washington, DC: National Center for Education Statistics, 1974.

Force, Ronald W. and Force, Jo Ellen, "Access to Alternative Catalogs: A Simulation Model", *College and Research Libraries*, 40:3 (May 1979), 234-239.

Franklin, Ralph W. and MacDonald, Clarice I., "The Washington Library Network", *Special Libraries*, 67:2 (February 1976), pp. 84-90.

Freedman, Maurice and Malinconico, S. Michael, *The Nature and Future of the Catalog: Proceedings of the ALA's Information Science and Automation Division's 1975 and 1977 Institutes on the Catalog*, Phoenix, Arizona: Oryx Press, 1979.

Furuya, Natsuko, ed., *What's In a Name? Control of Catalogue Records Through Automated Authority Files*, Toronto: University of Toronto, 1978.

Gore, Daniel, Kimbrough, Joseph, and Spyers- Duran, Peter, ed., *Requiem for the Card Catalog: Management Issues in Automated Cataloging*, Westport, Connecticut: Greenwood Press, 1979.

Haas, Warren J., Gwinn, Nancy E., and Jones, C. Lee, "Managing the Information Revolution: CLR's Bibliographic Service Development Program", *Library Journal*, 104:16 (September 15, 1979), pp. 1862-1870.

Hafter, Ruth, "The Performance of Card Catalogs: A Review of Research", *Library Research*, 1:3 (Fall 1979), pp. 199-222.

Hafter, Ruth, "Type of Search by Type of Library", *Information Processing and Management*, 15:5 (1979), pp. 261-264.

Hendricks, Donald D., *A Report on Library Networks* (Occasional Paper No. 108), Urbana, Illinois: University of Illinois Graduate School of Library Science, September 1973.

BIBLIOGRAPHY (Continued)

Hewitt, Joe A., "The Impact of OCLC", American Libraries, 7:5 (May 1976) pp. 168-275.

Hewitt, Joe A., OCLC: Impact and Use, Columbus, Ohio: Ohio State University Libraries, 1977.

Jacobs, Mary Ellen; Woods, Richard; and Yarborough, Judith, Online Resource Sharing II: A Comparison of OCLC, Incorporated, Research Libraries Information Network and Washington Library Network, California Library Authority for Systems and Services, January 1979.

Jones, C. Lee and Gwinn, Nancy, "Bibliographic Service Development: A New CLR Program", Journal of Library Automation, 12:2 (June 1979), pp. 116-124.

Kent, Allen and Galvin, T., ed., The Structure and Governance of Library Networks, New York: Marcel Dekker, 1979.

Kilgour, Frederick G., "Interlibrary Loans On-Line", Library Journal, 104:4 (February 15, 1979), pp. 460-463.

Knox, A. Whitney and Miller, Bruce A., "Predicting the Number of Public Computer Terminals Needed for an On-Line Catalog: A Queuing Theory Approach", paper presented at ALA Library Research Round Table Research Forum, Dallas, Texas, 1979.

Krikelas, James, "Catalog Use Studies and Their Implications" in Melvin Voigt, ed., Advances in Librarianship, Volume 3, New York: Seminar Press, 1972.

Lancaster, F. W., The Measurement and Evaluation of Library Services, New York: Information Resources Press, 1977.

Library of Congress, "Authority File Statistics Requirements" (NDO Nationwide Data Base Design Project Working Paper, Task 22.1), Washington, DC: Library of Congress, August 1978.

Library of Congress, "The Bibliographic Service Development Program of the Council on Library Resources", LC Information Bulletin, 38:14 (April 16, 1979), pp. 132-134.

Library of Congress, "Discussion Paper on the Incorporation of Series Authority and Treatment Data in the MARC Format" (NDO Nationwide Data Base Design Project Working Paper, Task 9), Washington, DC: Library of Congress, March 1979.

Library of Congress, "Retrospective Data in a Network Data Base" (Network Development Office, Nationwide Data Base Design Project Working Paper, Task 8), Washington, DC: Library of Congress, May 1979.

BIBLIOGRAPHY (Continued)

Lipetz, Ben-Ami, "Catalog Use in a Large Research Library", The Library Quarterly, 42:1 (January 1972), pp. 129-139.

Long, Philip L., Study of Message Text Formats: Bibliographic Search Queries (Network Planning Paper No. 5), Washington, DC: Library of Congress, 1979.

McCallum, Sally, Nationwide Data Base Design Project, Draft, Washington, DC: Library of Congress, 1979.

Malinconico, S. Michael, "Bibliographic Data Base Organization and Authority File Control", Wilson Library Bulletin, 54:1 (September 1979), pp. 35-45.

Malinconico, S. Michael and Fasana, Paul, The Future of the Catalog: The Library's Choices, White Plains, New York: Knowledge Industries Publications, 1979.

Maltby, A. and Sweeney, R., "The UK Catalogue Use Survey", Journal of Librarianship, 4:3 (July 1972), pp. 188-204.

Markey, Karen, Analytical Review of Catalog Use Studies, Columbus, Ohio: OCLC, Inc., February 1980.

Markuson, Barbara, "Ohio College Library Center System: A Study of Factors Affecting the Adaptation of Libraries to Online Networks", Library Technology Reports, 12:1 (January 1976).

Martin, Susan K., "An Odd Couple--AACR2 and Automation", American Libraries, 9:11 (December 1978), pp. 689-691.

Martin, Susan K., Library Networks 1978-79, White Plains, New York: Knowledge Industry Publications, 1979.

Matthews, Joseph R., "The Four Online Bibliographic Utilities: A Comparison", Library Technology Reports, 15:6 (November-December 1979), pp. 665-811.

Mowery, Judith, "Computer Terminals Required for Sufficient Access to an Online Public Catalog: A Queuing Theory Approach", unpublished paper, The Ohio State University Library, April 1978.

NCLIS/NBS Task Force on Computer Network Protocol, A Computer Network Protocol for Library and Information Science Applications, Washington, DC: National Commission on Libraries and Information Science, December 1977.

National Commission on Libraries and Information Science, Toward a National Program for Library and Information Services: Goals for Action, Washington, DC: NCLIS, 1975.

BIBLIOGRAPHY (Continued)

Network Technical Architecture Group, Message Delivery System for the National Library and Information Service Network: General Requirements (Network Planning Paper No. 4), Washington, DC: Library of Congress, 1978.

"Networking-North America", Bulletin of the American Society for Information Science, 5:5 (June 1979).

OCLC, Inc., Annual Report 1977-1978.

Online Review, 3:3 (September 1979), special issue on online cataloging.

Palmer, Richard P., Computerizing the Card Catalog in the University Library, Littleton, Colorado: Libraries Unlimited, 1972.

Palmour, Vernon E. and Roderer, Nancy K., "Library Resource Sharing Through Networks", in Annual Review of Information Science and Technology, Vol. 13, White Plains, New York: Knowledge Industry Publications, 1978.

Pierce, Anton R. and Taylor, Joe K., "A Model for Cost Comparison of Automated Cataloging Systems", Journal of Library Automation, 11:1 (March 1978), pp. 6-23.

Price, Derek de Solla, "A General Theory of Bibliometric and Other Cumulative Advantage Processing", Journal of the American Society for Information Science, 27:5 (September-October 1976), pp. 292-306.

Protocol Implementation Planners, Final Report to the Library of Congress Network Development Office on a Computer to Computer Communications Protocol Implementation Plan for the National Library and Information Service Network, New York: New York Public Library, August 1978.

Research Libraries Group, "Overview of the RLG Authority Subsystem: RLIN Functional Specifications", Research Libraries Group, June 1979.

Research Libraries Group, Progress Report June 1979.

Ross, Ryburn M., "Cost Analysis of Automation in Technical Services" in J. L. Divilbiss, ed., Proceedings of the 1976 Clinic on Library Applications of Data Processing: The Economics of Library Automation, Urbana, Illinois: University of Illinois Graduate School of Library Science, 1977.

Rouse, William, "Tutorial: Mathematical Modeling of Library Systems", Journal of the American Society for Information Science, 30:4 (July-August 1979), pp. 181-192.

Smith, Stanley V., Library Statistics of Colleges and Universities: Fall 1975 Institutional Data, National Center for Education Statistics, Washington, DC: Government Printing Office, 1977.

BIBLIOGRAPHY (Continued)

Socio-Economic Systems, Computer-Assisted Evaluation for Four Technical Processing System Bids: Final Report Prepared for the University of California Library System, Los Angeles, California: 1978.

Stevens, Charles, "Governance of Library Networks", Library Trends, 26:2 (Fall 1977), pp. 219-240.

Swanson, Don R., "Requirements Study for Future Catalogs", Library Quarterly, 42:3 (July 1972), pp. 302-315.

Tagliacozzo, Renata and Kochen, M., "Information-Seeking Behavior of Catalog Users", Information Storage and Retrieval, 6:5 (December 1970), pp. 363-381.

Thompson, James, "Ten Ways to Profit from a Long Engagement", American Libraries, 9:9 (October 1978), pp. 538-542.

Tracy, Juan I. and Remmerde, Barbara, "Availability of Machine-Readable Cataloging: Hit Rates for BALLOTS, BNA, OCLC, and WLN for the Eastern Washington University Library", Library Research, 1:3 (Fall 1979), pp. 227-281.

Waite, David P., "Basic Design Considerations for Library Networks", Network, 1:10 (October 1974), pp. 15-22.

Westat, Inc., Resources and Bibliographic Support for a Nationwide Library Program: Final Report to the National Commission for Libraries and Information Science, Washington, DC: Government Printing Office, August 1974.

Westat, Inc., A Study of the Characteristics, Costs, and Magnitude of Inter-Library Loans in Academic Libraries, Westport, Connecticut: Greenwood Publishing Company, 1972.

APPENDICES

APPENDIX 1

GLOSSARY

Definitions for terms marked with an asterisk (*) are from the NDO Network Planning Paper No. 2, "A Glossary for Library Networking", prepared by Dataflow Systems Inc., 1978.

Authority Control*

The functions involved in establishing, maintaining, and using authority files.

Authority File*

A set of records that identifies the established or authoritative forms for headings or access points for a set of bibliographic records. Authority files include cross references from variants to the preferred forms of headings, and links from earlier to later forms and between broader and narrower terms and related terms. An authority file may represent the authoritative forms of headings for an individual institution, a group of related institutions, or a network of related and/or unrelated institutions.

Authority Record*

A record of an individual heading in an authority file. An authority record may include heading variants, cross references to and from the headings, cataloging notes, historical information, and references to the source of a heading.

Batch Transmission

Electronic transmission of data after a number of similar input items have been accumulated and grouped together.

Bibliographic Record*

A collection of bibliographic data fields treated as one logical entity describing a specific bibliographic item.

Bibliographic Utility*

An organization that maintains online bibliographic data bases, enabling it to offer computer-based support to any interested users, including national library network participants.

CIP (Cataloging in Publication)

Cataloging done from galley sheets to be included in the final bibliographic item hence having no data in the collation fields. CIP is thus an incomplete record.

Catalog*

A set of bibliographic records, generally under control of authority files, which describes the resources of a collection, library, or network. It is the instrument by which bibliographic control is maintained for a collection, library, or network, and by which the relationship between individual bibliographic records may be indicated.

Cataloging Record*

A bibliographic record that describes a specific item and relates it to other items described in the file.

Copy Cataloging

See Shared Cataloging.

Current Title Cataloging

The process of creating catalog records for newly acquired items. See Retrospective Conversion.

Dedicated Line

A telecommunications circuit reserved for sole use. Dedicated lines are frequently lines leased from a common carrier.

Dial-Up Access

Use of the public telephone network to connect a terminal to a remote data base.

Distributed Data Bases*

Logically interconnected data bases or portions of data bases (indexes, locations, etc.) that reside in separate physical locations in a network.

Failed Search

A search of a local catalog which does not result in useful information.

Fill Rate

The ratio of interlibrary loan borrowing requests satisfied to the total number of interlibrary loan requests sent.

Hit Rate

The ratio of the number of successful searches of a data base to the total number of searches of that data base. See Verification Hit Rate and Holdings Hit Rate.

holdings hit rate: The ratio of ILL request titles that find a bibliographic record with holdings statements to the total number of items to be located.

verification hit rate: The ratio of ILL request titles that find a bibliographic record on a data base to the total number of requests to be verified.

Host Utility

Bibliographic utility with which the library under discussion is affiliated. See Linked Utility.

Host-to-Host Link

Telecommunication connection between the central computers in respective networks.

LC-MARC Record

Machine-readable cataloging provided by the Library of Congress as part of its MARC tape distribution service.

Link

An arrangement between two or more bibliographic utilities for the purpose of sharing bibliographic data.

Linked Utility

From the point of view of an individual library, a bibliographic utility whose resources are available only by means of a link with the library's host utility. See Host Utility.

Locations Data*

Data that identify the organization(s) holding one or more items described by a specific bibliographic record. Locations data do not include copy-specific data or site data for those items associated with an organization.

MARC Record

A machine-readable cataloging record. See LC-MARC Record.

Native Mode

Online interaction in which the search commands, response statements, and record formats used are those of the data base being accessed. If bibliographic utilities are linked, native mode refers to interactions in which a user of the host utility searches a linked data base by using the search commands of the linked utility and receives responses in the linked utility's format.

Online Access

Direct interaction with a central computer. Instructions or data are entered at a terminal, and the results of the computer processing are returned to the terminal.

Original Cataloging

Cataloging an item without using a bibliographic record produced by another library. See Copy Cataloging and Shared Cataloging.

Packet-Switching*

A type of data communication in which small defined blocks of data called packets are independently transmitted from point to point between source and destination, and re-assembled into proper sequence at the destination.

Protocol*

The conventions used in communicating between nodes and levels in a network, specifically a formal set of conventions governing the format and relative sequencing of message exchanges.

Reference Search

A search for information by a library patron or staff member. This report is most concerned with searches for bibliographic citations.

Research Libraries

For this report, research libraries are defined as members of the Association of Research Libraries (ARL).

Resource Sharing*

(1) Sharing of materials among cooperating libraries.

(2) Sharing of bibliographic data.

(3) Sharing of library processing services and the computer facilities, programs, and telecommunications to provide these services.

Retrospective Conversion

The process of creating machine-readable records for titles cataloged prior to the use of automated cataloging support at a given library. See Current Title Cataloging.

Shared Cataloging

Cataloging an item with the aid of a machine-readable record produced by another library. Also called Copy Cataloging. See Original Cataloging.

Union [illegible]alog*

A catalog that describes the contents of physically separate library collections, indicating by means of locations data the libraries in which a given item may be found.

Value-Added Network (VAN)*

A network operated by a private company that is authorized by a government agency to lease basic communications services from common carriers and specialized common carriers, to augment the services through additional facilities, such as switching centers and store-and-forward devices, and to resell the enhanced service to end users. Telenet and Tymnet are examples of value-added networks.

Verification

The process of confirming that the bibliographic citation for an item requested on interlibrary loan is complete and accurate.

APPENDIX 2A

THE IMPACT OF LINKING ON THE TIMING OF CURRENT TITLE CATALOGING

Figure 2A-1 is a graphic representation of the time-dependency of the benefit of linking for shared cataloging. When there is no link, the longer a library is willing to wait, the more original cataloging can be avoided. In most cases, virtually every title the library wishes to catalog will eventually appear on the host data base as new catalog records are added by other network libraries and as the data base is updated with LC-MARC records. Beyond a certain deadline, however, the benefit of using the machine-readable record is offset by the need to put the book on the shelf, and original cataloging is performed. As shown in Figure 2A-1, the earlier this deadline is imposed, the more original cataloging must be performed by the library. In short, the percentage of original cataloging that a library must perform for current titles decreases with time (Curve A).

When two or more data bases are linked, Curve A is displaced downward (Curve B). In other words, the added resources make records for some titles available earlier because they appear elsewhere before they appear on the host data base. If a cataloger has access to that other data, then after 2 weeks original cataloging is required for fewer titles than when no link is available. At any point on the time axis, linking will enable the library to save some of original cataloging effort, thereby avoiding some of the cost of cataloging. The maximum that can be avoided

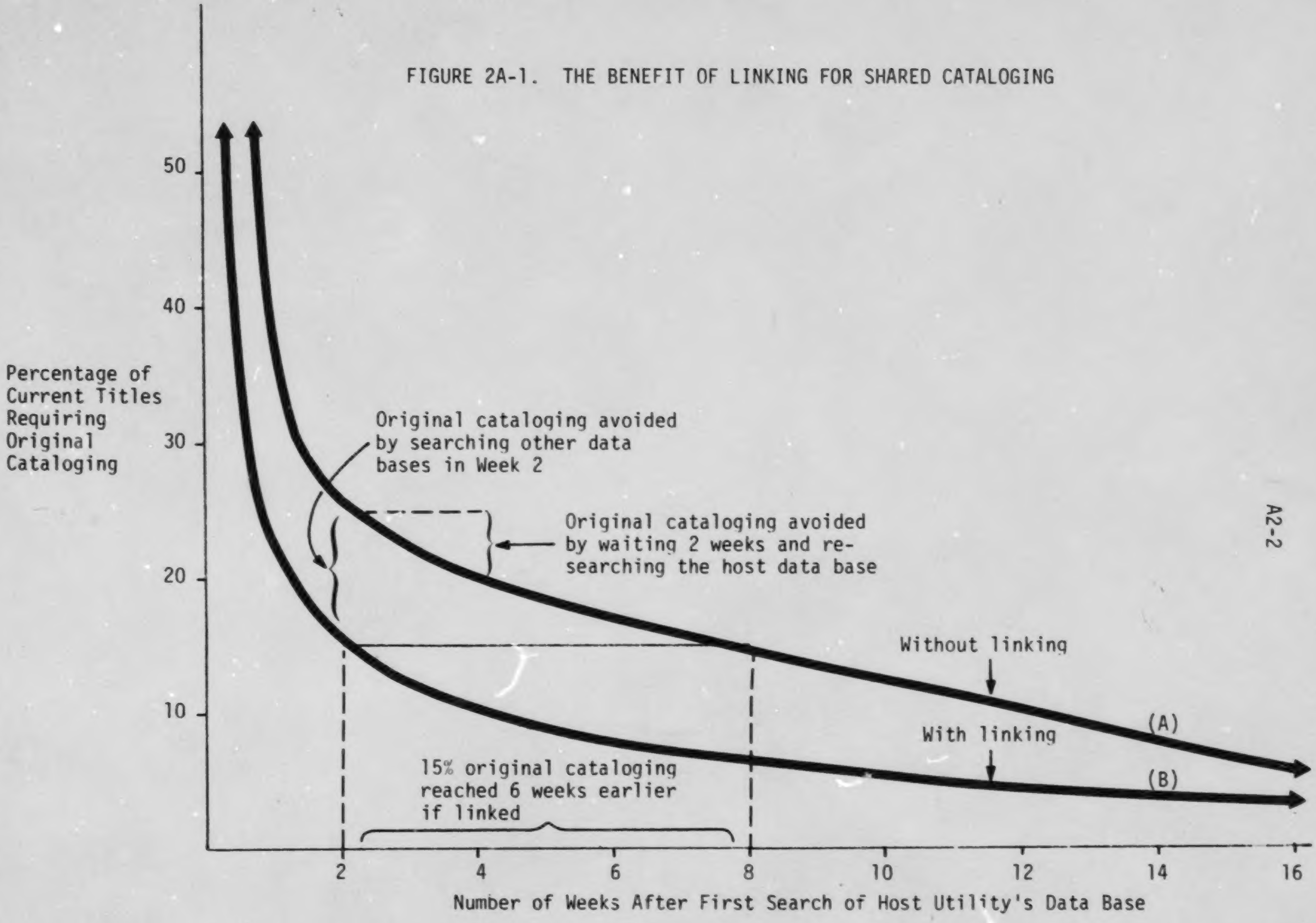


FIGURE 2A-1. THE BENEFIT OF LINKING FOR SHARED CATALOGING

out of a given number of titles is shown by the difference between the linked and unlinked curves at any point.

If the actual positions of the curves in Figure 2A-1 are known, then libraries may estimate the amount of cataloging cost they would avoid for a given number of titles if the resources of other data bases were available and were considered acceptable for copying. A library could then assess the tradeoff between waiting for a record to appear and cataloging the book quickly. Figure 2-12 in Chapter 2 of this report plots the curves for OCLC, RLIN, and WLN libraries, using the hit rates obtained in a test search of current titles.

APPENDIX 2B

COMPOSITION OF THE SAMPLE USED IN THE CURRENT TITLE TEST SEARCH

CURRENT TITLES SEARCHED, CATEGORIZED BY UTILITY OF CONTRIBUTING LIBRARY

Utility	Number of Titles	Percent of All Titles
OCLC	258	30
RLIN	176	21
WLN	173	21
(Weekly Record)	242	28
TOTAL	849	100

CURRENT TITLES SEARCHED, CATEGORIZED BY TYPE OF CONTRIBUTING LIBRARY

Library Type	Number of Titles	Percent of All Titles
ARL	212	25
Other Academic	159	19
Public	236	28
(Weekly Record)	242	28
TOTAL	849	100

CURRENT TITLES SEARCHED, CATEGORIZED BY TYPE OF MATERIAL

Type of Material	Number of Titles	Percent of All Titles
Fiction	87	10
Humanities	219	26
Social Science	219	26
Science and Technology	224	26
Non-English Language	100	12
TOTAL	849	100

APPENDIX 2C

WORKSHEET FOR CALCULATING ANNUAL CATALOGING COSTS AVOIDED BY LINKING

1. Current titles cataloged this year, per library
2. Reduction in percentage of titles requiring original cataloging
3. Potential number of titles for which original cataloging may be avoided this year
4. Fraction of these for which original cataloging is actually avoided
5. Number of titles for which original cataloging is actually avoided this year
6. Cost avoided per title if copy cataloging is done instead of original cataloging
7. Total cost of original cataloging avoided this year, per library
8. Cost avoidance per title cataloged (Line 7 ÷ Line 1)
9. Approximate number of affiliated institutions performing on-line cataloging
10. Aggregate cost avoidance for the network (Line 7 x Line 9)

APPENDIX 2D

AUTOMATED AUTHORITY FILES

Introduction

In addition to the benefits of shared cataloging records, linking may also prove to be useful for sharing of authority data. Several machine-readable authority files presently exist, including:

- Library of Congress (LC), the largest single producer of cataloging and authority records in the U.S., provides a large portion of recent, frequently used name authorities on magnetic tape. This includes authority records created in the last several years as well as names which occur 25 times or more on LC's machine-readable bibliographic file, a total of about 500,000 authority records. CLR is funding a project to convert an additional 104,000 LC name authority records to machine-readable form, especially those used from 3 to 24 times in LC's machine-readable files. About 450,000 subject headings are in machine-readable form for internal LC use only.

- OCLC now makes available online about 375,000 of the 500,000 LC name authority records that are now in machine-readable form. OCLC members may access this file to obtain authority information for use in online cataloging.

- The other two national libraries, the National Library of Medicine (NLM) and the National Agricultural Library (NAL) maintain automated or semi-automated authority files which are independent of, and different from, those of LC. The NLM file contains approximately 125,000 full-level name records and 84,000 provisional records; NAL has about 35,000 or more authority records. Discussions are being held that may lead to the sharing of authority work between the three national libraries.

- New York Public Library (NYPL) has developed an automated batch-processing authority control system which will become a major resource for the RLIN network.

- The Washington Library Network (WLN) maintains an authority control system for its users with a manual review and validation process performed at WLN for new authority records. The authority record format is not yet in conformance with LC format and a recently awarded CLR grant includes funding for conversion of WLN authority records to LC format.

- The University of Toronto Library Automated Systems (UTLAS) maintains an online authority control system.

- The National Library of Canada has an authority control subsystem within its automated DOBIS system.

In addition, automated authority files are maintained at some universities, including the University of Chicago and Northwestern University.

There are other automated authority systems under development. The Research Libraries Group is defining specifications for automated authority control using as base files the LC and NYPL authority files and accepting member-generated authority records under controlled conditions. In addition, RLIN and WLN are exploring methods of sharing authority data with one another.

The Library of Congress has already tested and is now discussing the further implementation of several cooperative authority file building ventures. These include a joint authority data base with NLM and NAL as mentioned above, corporate and personal name authority file cooperation with the Government Printing Office, a cooperative project in Africana with Northwestern University, and a project involving name authorities for Texas state government bodies with the Texas State Library. These projects will serve as prototypes for similar projects in other libraries.

Networks and Authority Control

Library networks are composed of autonomous institutions that may have different needs and practices, but some order of consistency in access points is required in order to produce a viable union catalog. Authority data used by individual network participants may differ in depth, content, or format:

- Depth--Libraries with large specialized collections may require a broader range of precise subject headings than libraries with more generalized collections. Northwestern University, for example, may have a larger set of subject headings in Africana than does LC. In that case, Africana catalogers will want to have access to this expanded set of subject headings before attempting to establish a new heading for their own collection. In addition, specialized libraries, such as law libraries, maintain more subject headings in their area of concentration than does LC.

- Content--There are many different authority files available at the national level--LC, NLM, National Library of Canada, Chemical Abstracts--and further authority records at local levels. Headings may conflict from file to file as authorities are established by catalogers in different milieus.

- Format--A MARC format for authority records was created through mutual agreement and compromise of many parties and is the format used to transmit LC authority records. Authority systems created prior to the development of the LC format (e.g., WLN authority file) are now faced with compatibility problems.

The bibliographic utilities must decide whether or not to enforce consistency upon their participants for the purpose of creating an internally consistent union catalog. WLN, for example, operates an authority control system to assure that all records at the network level are consistent. Deviations in entry are not tolerated. There is a capacity to support local additional information in specified fields. This information is only available for display to the contributing library. The plans for authority control at RLIN envision a similar flexibility.

In OCLC there is no enforcement of authority control except where an LC-MARC record (subject to LC authority control) replaces member-generated cataloging. This occurs whenever a match can be identified between a previously entered member record and a new LC-MARC record. Even in this case, however, the matching algorithm may not catch what is in reality a duplicate record because the authority tolerance in the algorithm is so wide. Therefore, two or even three records for the same bibliographic item may sometimes exist concurrently on the OCLC data base. To some extent, the problem may be reduced for name headings with the availability of LC's name authority file for online searching at OCLC, but there will still not be any automatic checking or validation to assure internal consistency in the OCLC union catalog.

Summary of Authority Capabilities of the Bibliographic Utilities

The four bibliographic utilities which are the focus of this study have different authority capabilities:

- LC performs high-quality, single library authority control internally and, by extension, provides authority control over the LC-MARC records which carry headings established in accordance with LC authority control practices.

- RLIN does not presently support authority control but is planning an authority control subsystem based on different levels of control--the network level and the local library level--and different hierarchies of record modification capability.

- WLN performs automated authority control with manual validation of new authority records.

- OCLC has loaded the LC name authority file for online search by its members. No automated authority control is yet available to OCLC members, but the need for authority control is being explored. The decision to implement will be affected by cost of implementation, technical factors such as resource allocation, user work flow considerations, and user needs. The capability to automatically create required cross-references and to automatically change a given heading for all records where that heading appears is not easily compatible with the data base structure at OCLC. The technical feasibility of authority control on the system has recently been a matter of great concern.

The Potential Benefits of Linking Authority Files

Given that all utilities will have equal access to LC's machine-readable name authority records, what benefits are possible from linking authority files between utilities?

- Cost Avoidance Benefits--A major proportion of professional staff time in original cataloging is devoted to authority work. If shared authority data can reduce the staff time required to create new headings beyond those already supplied by LC, then some of the cost of authority work may be avoided.

- Increased Coverage--Subject division headings, for example, are searchable on WLN. Online access to this file for searching would add not only additional records but an additional type of authority record for use by OCLC users. The developing authority control system at RLIN will provide a similar capability.

- Increased Consistency at the National Level--The increased sharing of authority data will combine with the increasing tendency to use an existing record rather than create a new one which should result in a tendency toward consistency in headings. This will promote easier exchange of bibliographic records among utilities and produce a higher level of recall confidence in a bibliographic search.

- Quality of Authority Records--A cooperative effort seems to be required to create a system at the national level to produce and share high quality authority records. For example, cooperation between

LC, with its commitment to high-quality authority records, and certain large research libraries, with their respective subject authority competencies, may create a national network for authority work. Other libraries possessing specialized knowledge in certain areas (e.g., local place names, government, corporate names) will increase the coverage and precision of a cooperative authority system. Such a cooperative effort may lead to the sharing of bibliographic records as well as authority records.

The Cost of Establishing Authority

Very little has been published on the cost of authority work. Directors or department heads at the Library of Congress, the National Library of Medicine, and several major research libraries were interviewed to determine cost figures calculated at these institutions. All institutions contacted give considerable importance to thorough authority work and are known for a high level of performance in this area. Consequently, costs reported here may be higher than those characteristic of the library community at large. They will, however, have relevance for the research library community for which authority control is a particularly important benefit to be derived from using a bibliographic utility.

Ritvars Bregzis of the University of Toronto has published data indicating that establishing main entries accounts for 35 percent of the cost of cataloging*. Subject authority and added entry work would incur additional costs. These were not specifically enumerated, though subject cataloging accounted for 20 percent of cataloging costs. These figures include catalog maintenance, while figures reported by other libraries generally include only the cost of cataloging a given item.

The costs reported by the librarians interviewed varied widely, since cost studies at each library reflected conditions unique to that library. Since figures are not strictly comparable, individual reports are given below in narrative form:

- Average costs of $65 to $75 per title for complete original cataloging, including input into the local computer. This is staff labor cost only, including descriptive and subject cataloging, establishment of entry, shelf listing, record input, and supervision costs. Of these costs, $10 to $20 was attributed to the establishment of main entry (name authority work only). An additional $5 to $6 is attributable to subject authority work. Hence $15 to $26 per title may be attributed to authority work, or about 20 to 40 percent of the cost of cataloging.

---

* Ritvars, Bregzis, "Integrated Access to Information and the Economic Reality", in Natsuko Furuya, ed., What's in a Name: Control of Catalog Records Through Automated Authority Files, University of Toronto, 1978.

- Average costs of $65 per title for complete original cataloging, $39 per title (60 percent of cataloging costs) for name authority work, including main entry and personal or corporate name added entries. This includes entry into an online authority file with online update and validation functions.

- Cost of cataloging is $10 per title for titles found on OCLC and $15 to $30 for titles requiring original cataloging. Costs of authority work are reported on a per heading basis for name headings only at $7.50 per heading, a cost comprised of 45 minutes worth of non-professional search time and 15 minutes of professional verification time per heading. This means that for a title found on OCLC and requiring establishment of one name heading, the name authority work would constitute 43 percent of the cost of cataloging ($7.50 divided by $17.50). For a title with two name headings, authority work would take up to 60 percent of the cataloging cost ($15 divided by $25). The percentage of original cataloging cost incurred by authority work for original cataloging

would be 20 to 33 percent for a single name heading,
33 to 50 percent for two headings.

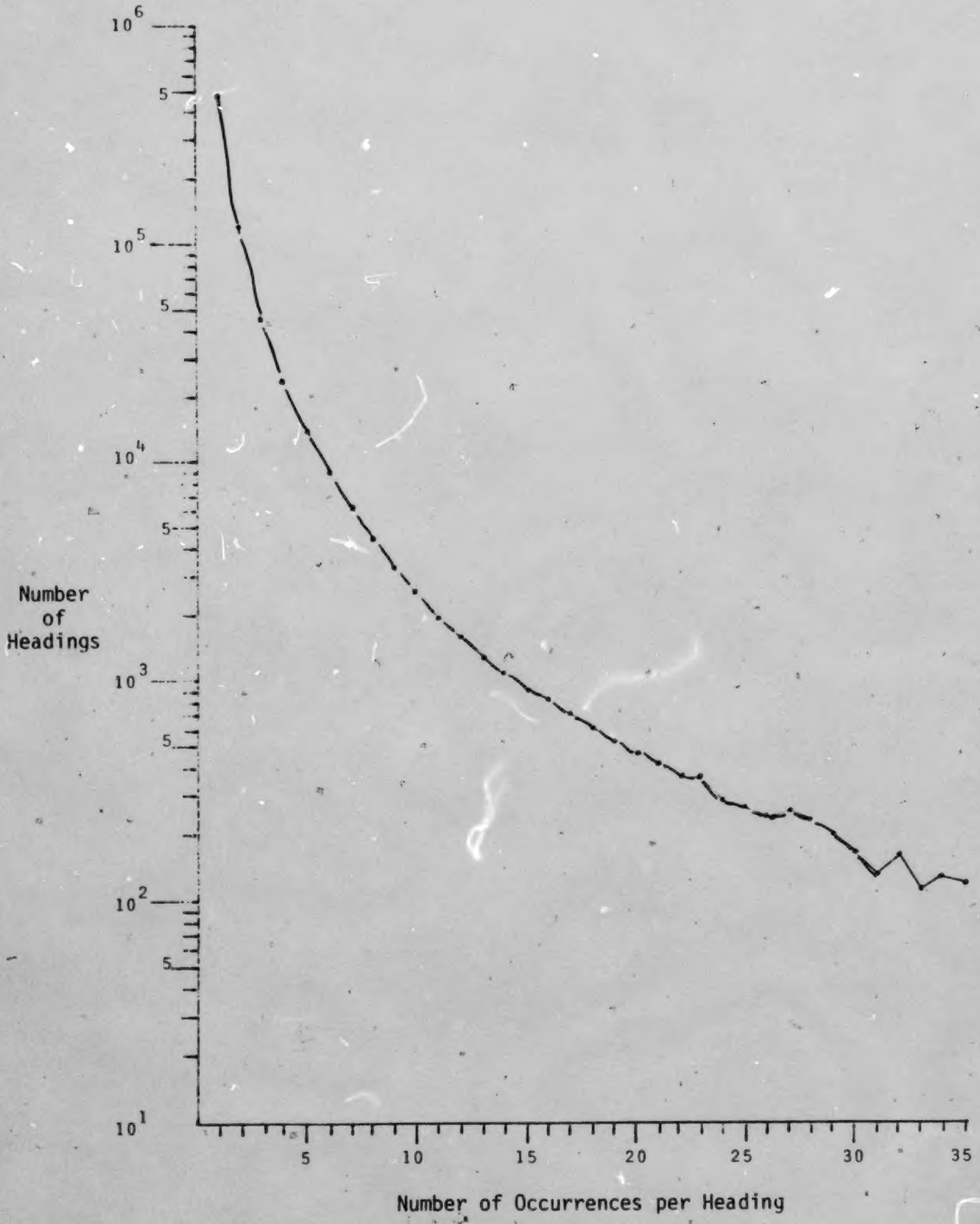
These figures demonstrate the potential for cost savings if an online authority file were available. Total savings will depend, of course, on the total volume of cataloging and authority work, and this volume will vary from library to library.

As was the case with bibliographic records, not all of these name headings are necessarily candidates for a link. About two-thirds of the unique name headings in the LC file are used only once. An additional 17 percent are used only twice (see Figure 2D-1). We may conclude that only the remaining 16 percent are likely to receive extended use, and these headings, then, are candidates for a link. In fact, this 16 percent represent those records being converted to machine-readable form through the CLR grant mentioned above. These records are to be distributed on tape to all the utilities as part of the LC authority record distribution service.

Implementing an Authority Link

There are several ways to implement authority linkages. One way is a simple tape exchange of authority records such as LC presently performs with name authorities. This distribution would involve only the approximately 16 percent of the authority records that are used most often. Such a batch system, however, fails to cope with the dynamic nature of authority files and may lead to problems of redundancy and inconsistency through creation of duplicate records between updates. An online access

FIGURE 2D-1. AUTHORITY HEADING OCCURRENCES--LC NAME AUTHORITY FILE



link such as is being considered for other applications is particularly favored for authority work where file update must be rapid.

Linking structure presents different problems, but steps are already being taken to implement certain links. A program to develop an authority link between WLN and RLG is already underway. The organizational framework already exists within which this will take place. The funding process has begun for upgrading WLN authority records to LC format.

A general plan for implementing a nationwide authority file has been outlined in a white paper prepared by the Council on Library Resources and entitled An Integrated Consistent Authority File Service for Nationwide Use (December 4, 1979). Developed through discussions with the bibliographic networks and the national libraries, the concept calls for the following:

- Agreement on the possibility and desirability of designing, establishing and maintaining an integrated, consistent, nationwide authority file.

- Large general research libraries serving as contributing institutions using the LC name authority file as the base for development, perhaps merged with the files of NYPL, WLN, Stanford, or Chicago.

- LC responsibility for maintenance and input of authority records, training staff at contributing institutions, and quality control.

- RLG/RLIN responsibility for physical housing, file management, programming, display formats, and output production.

Battelle interviewed a number of knowledgeable individuals with experience and expertise in automated authority files and authority control. Many of the concepts in the CLR paper were mirrored in remarks made during these interviews, especially the desire for a consistent nationwide authority file. These comments indicate the practicability of the CLR plan.

Some other points of view expressed during these discussions included:

- The establishment of "centers of excellence", or subject specialities among libraries contributing to the national file. The problem is raised, then, of overlap in coverage and conflicting headings between two libraries, each of which has reason to consider the heading under its subject purview.

- Establishment of several levels of authentication; for example, the national libraries as the top level, a network level required to produce records accepted as

> authentic across an entire network, and local level visible only to the local library. However, different authorities available at different levels of a hierarchy would raise serious problems of consistency in offline catalogs produced from a data base.

While opinions may differ about the most appropriate means for achieving a consistent nationwide authority data base, there is agreement among many experts that such a data base would be of benefit in terms of reducing the amount of authority work required in libraries as well as improving precision in bibliographic searching. The bibliographic utilities may play a crucial role, since they have been major forces in the development to date of automated authority files.

APPENDIX 3A

THE EFFECTS OF UTILITY USE ON ILL COSTS

A summary of the comments of seven ILL librarians
interviewed in December 1979 and January 1980.

Effects on Lending Costs

Verification:	Great reduction in time needed for those few requests that must be verified. Little time savings overall.
Regrets Notices:	Any increase in this activity is probably not due to utility usage, since very little of the increase in request volume is attributable to utility membership.
Net Effect:	No change in the unit cost of lending.
Total Cost Trend:	Higher costs due to inflation and higher volume of incoming requests, with no reduction in unit cost to offset the increase.

Effects on Borrowing Costs

Verification:	Great reduction in time needed per request.
Location:	Moderate reduction in time needed per request.
Selection:	Slight increase in time needed due to unfamiliarity with codes and sheer number of holdings for some items.
Re-Issuing:	Moderate reduction in time spent re-issuing requests because of higher first-time fill rate.
Pursuit of Unfilled Requests:	Slight increase in time taken because availability of holdings information makes it possible to pursue a search further than before.
Net Effect:	Significant reduction in the unit cost of borrowing.
Total Cost Trend:	Reduction in unit cost offsets much of the increase in total costs that would normally result from inflation and higher borrowing volume.

Effects on Staffing and Training

Staffing:	No change in salary expenses.
Training:	Slight reduction in time needed.
Net Effect:	No impact on costs since any time saved is usually spent on other ILL activities or on handling greater volume.

Aggregate Effect on ILL Costs

Net Effect:	Moderate reduction in cost per ILL transaction.
Total Cost Trend:	Greater productivity due to use of utility could result in reduced cost of ILL. In reality, ILL budgets are usually not reduced, but rather they continue to rise with inflation and higher request volume. The benefit of the utility is that the budget increase can be smaller than it would otherwise have been because more work can be done with the same staff.

APPENDIX 3B

ILL REQUEST PATTERNS

Chapter 3 discusses the hit rates encountered in a test search of 807 monograph borrowing requests in February 1980. The results of the test search suggest a general pattern of ILL borrowing.

Patrons generally request to borrow books that are "widely held" among OCLC participants but "not widely held" among RLIN or WLN libraries. Many definitions of "widely held" may be proposed, and the choice is somewhat arbitrary. In this study, an item is considered widely held among OCLC libraries if more than five holding symbols appear on its record. An item is widely held among RLIN and WLN libraries if more than two libraries hold the book in each case. The great difference in the size of memberships required that the cut-off point used for WLN and RLIN be lower than that for OCLC.

Figure 3B-1 shows how the results of the test search were distributed among three categories for each utility: widely held, not widely held, and no record. Fifty-seven percent of the titles searched were held by more than 5 OCLC libraries, 21 percent of the requests were for books that were widely held by WLN users, and only 13 percent were for items held by more than 2 RLIN libraries.

Figure 3B-2 gives the percentages of requests for items that were not widely held by members of one utility and were widely held by members of another. For example, 17 percent of the titles searched were not widely held among RLIN users but were widely held by OCLC libraries. Only one request was for a book that was not widely held in OCLC but was widely held elsewhere.

Figure 3B-3 shows that this pattern also holds when the sample is broken down by the utility affiliation of the libraries that contributed requests to the study.

FIGURE 3B-1. PERCENTAGE OF REQUESTS FOR BOOKS WIDELY HELD* IN EACH UTILITY

Results of a test search of 807 monograph borrowing requests in February 1980.

	OCLC	RLIN	WLN
Widely Held	57%	13%	21%
Not Widely Held	20%	20%	12%
No record	23%	67%	67%
	100%	100%	100%

* OCLC: widely held if more than 5 holding symbols
RLIN: widely held if more than 2 libraries hold
WLN : widely held if more than 2 libraries hold

FIGURE 3B-2. PERCENTAGE OF REQUESTS FOR BOOKS NOT WIDELY HELD IN ONE UTILITY BUT WIDELY HELD* IN ANOTHER

Results of a test search of 807 monograph borrowing requests in February 1980.

	Widely Held in OCLC	Widely Held in RLIN	Widely Held in WLN
Not Widely Held in OCLC	-	0	0.12%
Not Widely Held in RLIN	17%	-	8%
Not Widely Held in WLN	12%	2%	-

*OCLC: widely held if more than 5 holding symbols
RLIN: widely held if more than 2 libraries hold
WLN : widely held if more than 2 libraries hold

FIGURE 3B-3. PERCENTAGE OF REQUESTS FOR BOOKS NOT WIDELY HELD IN ONE UTILITY BUT WIDELY HELD* IN ANOTHER

Requests broken down by utility affiliation of contributing library.

Results of a test search of 807 monograph borrowing requests in February 1980.

	Widely Held in OCLC	Widely Held in RLIN	Widely Held in WLN
Not Widely Held in OCLC (n = 674)	-	0	0.15%
Not Widely Held in RLIN (n = 82)	52%	-	23%
Not Widely Held in WLN (n = 51)	35%	7%	-

* OCLC: widely held if more than 5 holding symbols
RLIN: widely held if more than 2 libraries hold
WLN : widely held if more than 2 libraries hold

APPENDIX 3C

COMPOSITION OF THE SAMPLE USED IN THE MONOGRAPH BORROWING REQUEST TEST SEARCH

BORROWING REQUESTS SEARCHED, CATEGORIZED BY TYPE OF CONTRIBUTING LIBRARY

Library Type	Number of Titles	Percent of All Titles
Research Library Title	285	35.3%
Academic Library Title	345	42.7%
Public Library Title	177	22.0%
TOTAL	807	100.0%

BORROWING REQUESTS SEARCHED, CATEGORIZED BY TYPE OF MATERIAL

Type of Material	Number of Titles	Percent of All Titles
Social Science	184	22.8%
Science/Technology	122	15.1%
Biography, History, Travel	107	13.3%
Foreign Language	122	15.1%
Other	272	33.7%
TOTAL	807	100.0%

BORROWING REQUESTS SEARCHED, CATEGORIZED BY PUBLICATION DATE

Publication Date	Number of Titles	Percent of All Titles
1975-1979	227	28.1%
1960-1974	271	33.6%
1900-1959	179	22.2%
1800-1899	50	6.2%
No Date	80	9.9%
TOTAL	807	100.0%

BORROWING REQUESTS SEARCHED, CATEGORIZED BY UTILITY

Utility	Number of Titles	Percent of All Titles
OCLC	674	83.5%
RLIN	82	10.2%
WLN	51	6.3%
TOTAL	807	100.0%

APPENDIX 3D

ANALYSIS OF REQUESTS FOR WHICH NO HOLDINGS WERE FOUND

Results of a test search of 807 monograph
requests in February 1980.

Category	Number of Requests in Category	Number of No-Hits in Category	Percentage No-Hits in Category	Percentage of All No-Hits (n=169)
Type of Material				
Social Science	184	24	13.0%	14.2%
Science/Technology	122	32	26.2%	18.9%
Biography, History, Travel	107	16	14.9%	9.5%
Foreign Language	122	61	50.0%	36.1%
Other	269	34	12.6%	20.1%
Not Given	3	2	66.7%	1.2%
TOTAL	807	169		100.0%
Date of Publication				
1975-1979	227	18	7.9%	10.7%
1960-1974	271	56	20.6%	33.1%
1900-1959	179	60	33.5%	35.5%
1800-1899	50	17	34.0%	10.0%
No Date	80	18	22.5%	10.7%
TOTAL	807	169		100.0%
Type of Library				
Research	285	93	32.6%	55.0%
Academic	345	62	17.9%	36.7%
Public	177	14	7.9%	8.3%
TOTAL	807	169		100.0%
Utility				
OCLC	674	147	21.8%	87.0%
RLIN	82	16	19.5%	9.5%
WLN	51	6	11.7%	3.5%
TOTAL	807	169		100.0%

APPENDIX 3E

WORKSHEET FOR CALCULATING THE VERIFICATION COSTS AVOIDED BY LINKING

1. Number of monograph borrowing requests processed in a year
2. Reduction in percentage of titles that cannot be verified online
3. Number of requests which can be verified online only if linked (Line 1 x Line 2)
4. Cost avoided per request if on-line verification is possible
5. Total cost of verification that can be avoided in a year if linked (Line 3 x Line 4)
6. Cost avoided per request processed (Line 5 ÷ Line 1)
7. Approximate number of libraries (public, research, and other academic only)
8. Aggregate cost avoided in a year by all affiliated libraries (Line 5 x Line 7)

APPENDIX 3E (Continued)

WORKSHEET FOR CALCULATING THE LOCATION COSTS AVOIDED BY LINKING

1. Number of monograph borrowing requests processed in a year
2. Reduction in percentage of titles that cannot be located online
3. Number of requests which can be located online only if linked (Line 1 x Line 2)
4. Cost avoided per request if online location is possible
5. Total cost of location that can be avoided in a year if linked (Line 3 x Line 4)
6. Cost avoided per request processed (Line 5 ÷ Line 1)
7. Approximate number of libraries (public, research, and other academic only)
8. Aggregate cost avoided in a year by all affiliated libraries (Line 5 x Line 7)

APPENDIX 3F

REVIEW OF LITERATURE ON THE COST OF ILL

The ILL cost studies performed in the past 10 years exhibit a great deal of variability in focus, methodology, and results. Some studies are based on a meticulous task analysis of ILL labor activities, while others use fairly broad categories for collecting time data. Some measure all direct costs, some measure only direct labor. A few include overhead in the final cost total, but there appears to be no standard method for computing overhead costs. The amount of detail concerning photocopying and delivery costs varies substantially among the studies. Most studies attempt to isolate costs for lending and borrowing, but a few measure only the total costs attributable to the ILL function. Some identify the costs associated with filled and unfilled requests for both borrowing and lending. Almost all of the studies analyze costs for academic libraries only.

Figure 3F-1 summarizes the results of several published studies of the unit cost of ILL. Recent data were also provided by three research libraries to augment the data found in the reports and articles that were reviewed. The costs reported in each case have been adjusted for inflation.

The average cost of a filled lending request is about $8.75. The average cost of a filled borrowing request is about $14.00.

FIGURE 3F-1. RESULTS OF LITERATURE REVIEW ON UNIT COSTS OF ILL

SOURCE	PUB DATE	DATE of DATA	INFLATION FACTOR[1]	LENDING				BORROWING				OVERHEAD FACTOR USED
				FILLED		RECEIVED		FILLED		SENT		
				REPORTED	ADJUSTED	REPORTED	ADJUSTED	REPORTED	A DJUSTED	REPORTED	ADJUSTED	
Research Library	--	1979	1.000	$ 5.75	$ 5.75			$ 8.64	$ 8.64			65% of salaries
Research Library	--	1978	1.130	9.00	10.17			11.00	12.43			50% of direct labor cost (from Westat)
Research Library	--	1978	1.130	14.58	16.48							47% of salaries and benefits
King[2]	1979	1978	1.130	8.40	9.49			11.60	13.11			not given
Bellanti and Widdifield	1978	1976	1.294			(2.40)[3]	3.11					50% of direct labor cost (not used by author)
Neustaetter	1977	1976	1.294			2.99	3.87					170.8% of total direct costs
Baumol and Ordover	1976	1975	1.384					12.26	16.97			
West and Koch	1974	1973	1.631			2.68	4.37			6.33	10.32	50% of direct labor cost (from Westat)
Cossar	1973	1972	1.705			(2.91)	4.96			(6.12)	10.43	50% of direct labor cost (not used by author)
Westat: main libraries	1972	1971	1.780	5.82	10.36			7.61	13.55			50% of direct labor cost
Westat: branch libraries	1972	1971	1.780	3.74	6.66			11.65	20.74			50% of direct labor cost
Llewellyn and Allen	1971	1970	1.884	(3.57)	6.73			(7.07)	13.32			50% of direct labor cost (not used by authors)
Spencer	1971	1969	1.981	2.36	4.68	1.53	3.03					45% of salaries
Mean Costs					$8.79		$3.87		$14.11		$10.38	
Standard Deviation					3.76		0.82		3.80		.078	

[1] See Figure 3F-2.

[2] King reports "typical" costs derived from data presented in three other studies. They are listed as items 4, 6, and 9 in the bibliography.

[3] A unit cost figure in parentheses on the chart includes overhead cost that was not included by the author of the study in question. Fifty percent overhead on direct labor cost has been added to make these data comparable with data from other studies. This overhead rate is recommended in the 1972 Westat report and has been used in many subsequent studies.

The numbers in parentheses were derived in the following manner:

(1) (Reported unit cost) - ($.60 indirect and non-labor direct costs) = direct labor cost

(2) (Direct labor cost) x (1.5) = direct labor cost plus 50 percent overhead

(3) (Direct labor cost plus overhead) + ($.60 indirect and non-labor direct costs) = unit cost including overhead.

FIGURE 3F-2. INFLATION FACTORS USED TO ADJUST COSTS REPORTED IN PREVIOUS YEARS

Year	Implicit Price Deflator(1)	Inflation Factor(2)
1969	86.1	1.980
1970	90.5	1.884
1971	95.8	1.780
1972	100.0	1.705
1973	104.6	1.630
1974	112.7	1.513
1975	123.2	1.384
1976	131.8	1.294
1977	140.7	1.212
1978	150.9	1.130
1979	(170.5)(3)	1.000

(1) Source: Survey of Current Business, U.S. Department of Commerce, Bureau of Economic Analysis, Washington, DC, 1979.

(2) $\frac{\text{1979 implicit price deflator}}{\text{19XX implicit price deflator}}$

(3) 1979 implicit price deflator has not yet been published. Using an assumed 1979 inflation rate of 13%, the 1979 deflator may be calculated as follows:

(1978 deflator) x (1 + 1979 inflation rate)

150.9 x 1.130 = 170.5

References

(1) Baumol, W. J. and Ordover, J. A., "Public Good Properties in Reality: The Case of Scientific Journals", in Susan K. Martin, comp., *Information Politics: Proceedings of the American Society for Information Science,* 39th Annual Meeting, Washington, D.C., ASIS, Vol. 13, 1976, pp. 442-470.

(2) Bellanti, R. and Widdifield, C., Interlibrary Lending Cost Analysis: UCLA Biomedical Library: Preliminary report. Paper presented at the Annual Meeting of the Medical Library Association, Chicago, March 1978, ED 163 927.

(3) Brown, M. K. and McHugh, A. L., *Survey of Costs in Technical Processing and Interlibrary Loan*, Western Interstate Library Coordinating Organization (WILCO), Boulder, Colorado, December 1976.

(4) Cossar, Bruce, "Interlibrary Loan Costs", *RQ*, Spring, 1973, pp. 243-246.

(5) King, Donald W., "Pricing Policies in Academic Libraries", *Library Trends*, Summer 1979, pp. 47-62.

(6) King, Donald W., et al., "The Journal in Scientific Communication", Rockville, Maryland, King Research, Inc., 1978 (NSF Contract No. DSI 75-06942).

(7) Llewellyn, G. M. and Allen, K. S., "Interlibrary Loan Costing", *PNLA Quarterly*, Summer 1971, pp. 9-15.

(8) Neustaetter, T. M., Direct and Overhead Interlibrary Lending Costs of the Pacific Southwest Regional Medical Library Service, 1 June 1976-31 May 1977. Master's thesis, University of California at Los Angeles, 1977, ED 145 813.

(9) Spencer, C. C., "Random Time Sampling with Self-Observation for Library Cost Studies: Unit Costs of Interlibrary Loans and Photocopies at a Regional Medical Library", *Journal of the American Society for Information Science*, May-June 1971, pp. 153-160.

(10) West, M. W. and Koch, R. A., Interlibrary Request and Loan Study: 1973 report. San Jose State University, Department of Librarianship, 1974, ED 104 423.

(11) Westat, Inc., *A Study of the Characteristics, Costs, and Magnitude of Interlibrary Loans in Academic Libraries*, Westport, Connecticut: Greenwood Publishing Company, 1972.

APPENDIX 4A

ONLINE CATALOG REQUIREMENTS

As part of the investigation of linking for online reference, Battelle prepared an estimate of the total traffic that might be expected to use an online reference service, and based on this, projected the number of terminals and computer resources that would be required to support such a service. A "worst case" scenario was adopted for the estimation of terminals and computer resources in order to reflect peak load traffic in the results. The investigation is discussed under the following components:

- Total catalog traffic
- Required terminals
- Required computer resources

Total Catalog Traffic

For the purposes of this task, total potential online reference searching roughly equals the present use of card, COM and book catalogs. The initial problem, then, was to estimate present catalog use traffic in the libraries using one of the utilities.

Lancaster (1) and Hafter (2), have suggested that only a handful of past studies of the card catalog were methodologically sound. Among these were Lipetz, Palmer, Tagliacozzo and Kochen, and Swanson (3). These studies were useful for determining how the card catalog is used, particularly in academic libraries. The Lipetz study was especially valuable because it indicated not only how the card catalog is used, but what the traffic pattern and volume was at that time at Yale. Our attention quickly focused on two key concepts by Lipetz: his catalog traffic curve by hour of the day and his observation that there is a strong one-to-one correlation between library book circulation and card catalog use. In other words, the card catalog was used once for each book circulated, and therefore, one variable can be used to predict the other in academic libraries. Krikelas (4) confirmed this observation using Palmer's data. This observation is particularly important since catalog use data is so sparse, but circulation data is relatively easy to obtain.

For purposes of a gross estimate of card catalog traffic, and because Lipetz is recognized as the methodological standard in this field (2), it was decided to use his findings to typify catalog activity in all academic libraries. This was, of course, an over-simplification, but sufficiently trustworthy data were not available to accurately characterize academic libraries.

There is no study comparable to Lipetz' in the area of public libraries which presents an hourly traffic curve as well as a reasonable correlation between catalog use and circulation. The closest approximation is a recent study of COM catalog use in the Los Angeles County Library performed by Brett Butler et al (5). This study showed hourly traffic patterns at the catalog for five categories of test libraries, a total of 18 sites. The five traffic curves developed by Butler were combined into a single composite curve to typify the pattern of public library catalog use.

It was felt that this composite curve was a reasonable representation of card catalog use in public libraries since it combined traffic at small branches with that of intermediate and large branches.

A correlation between catalog use and circulation is difficult to estimate for public libraries short of undertaking a study on a scale similar to that of Lipetz' work. Butler's data were used to develop the general shape of a catalog traffic curve for public libraries. In addition, direct observations were made of a catalog at the main library and a typical branch in a medium-sized metropolitan library system. These observations were normalized by a factor based upon the composite curve developed from Butler's data to derive an average hourly catalog use figure. Then, this figure was compared to circulation figures to produce the desired correlation. The overall correlation was calculated at about 1:8; that is, for every catalog use, 8 items are circulated.

There is no question that there is considerable variance in the proportion of the use of the card catalog to the number of items circulated depending on the size of the library. Smaller academic libraries may tend to shade toward the public library correlation figure while larger public libraries may approach the academic library correlation figure. The complexity of the question is recognized, however. For the purpose of this study, our simplifications were both necessary and deemed sufficient.

The following procedure was used to estimate card catalog use:

(1) Circulation data were compiled for each academic and public library belonging to RLIN and WLN; for

OCLC, academic and public library circulations were estimated via a random sample of the network because of its large size. Five strata were defined for the OCLC network to ensure better sampling. They were: research libraries (ARL members), large academic (collections greater than 100,000 volumes), small academic (collections smaller than 100,000 volumes), large public (over 100,000 volumes) and small public (under 100,000 volumes) libraries.

(2) All libraries, academic and public, were grouped by time zone divisions for each utility. Circulation data were totaled according to the time zone divisions (Figure 4A-1).

In the case of OCLC, the first step was to estimate an average annual circulation per library for each of the five strata using the random sample selected for each stratum (see upper half of Figure 4A-1, OCLC sample). Next, circulation was estimated for the OCLC libraries within each time zone by multiplying the average circulation for each stratum by the actual number of OCLC libraries in each stratum time zone combination (see upper half of Figure 4A-1). Finally, OCLC strata 1 thru 3

FIGURE 4A-1. ANNUAL AND DAILY LIBRARY CIRCULATION DATA BY UTILITY BY TIME ZONE

		OCLC Sample		EST		CST		MST		PST	
	Network	N	Ave. Circ. This Category	# Libraries	Total Circ.	# Libraries	Total Circ.	# Libraries	Total Circ.	# Libraries	Total Circ.
(1) Research Libraries	OCLC	10	705,436	42	29,628,312	20	14,108,720	4	2,821,744	2	1,410,872
(2) Large Academic Libs.	OCLC	72	114,324	317	36,240,708	157	17,948,868	18	2,057,832	34	3,887,016
(3) Small Academic Libs.	OCLC	6	46,500	210	9,765,000	138	6,417,000	13	604,500	8	372,000
(4) Large Public Libs.	OCLC	26	1,831,862	50	91,593,100	26	47,628,412	3	5,495,586	3	5,495,586
(5) Small Public Libs.	OCLC	7	345,848	49	16,946,552	35	12,104,680	5	1,729,240	3	1,037,544
Academic Libraries 1 + 2 + 3	OCLC	Circ/Yr Circ/Day (7-Day Wk)			75,634,020 210,094		38,474,588 106,874		5,484,076 15,234		5,669,888 15,750
Public Libraries 4 + 5	OCLC	Circ/Yr Circ/Day (6-Day Wk)			108,539,652 346,772		59,733,092 190,841		7,224,826 23,083		6,533,130 20,873
Academic Libraries	RLIN	Circ/Yr Circ/Day(7-Day Wk)			2,932,900 8,147		-- --		1,033,383 2,871		7,999,060 22,220
Public Libraries	RLIN	Circ/Yr Circ/Day (6-Day Wk)			9,130,268 29,170		739,988 2,364		168,279 538		30,613,154 97,806
Academic Libraries	WLN	Circ/Yr Circ/day (7-Day Wk)			-- --		-- --		612,967 1,703		3,128,056 8,689
Public Libraries	WLN	Circ/Yr Circ/day (6-Day Wk)			-- --		-- --		748,379 2,391		15,316,568 48,935
Public Libraries ,	WLN	Circ/Yr Circ/Day (6-Day Wk)			AHST* 606,999 1,939	Academic Libraries, WLN			Circ/Yr Circ/Day (7-Day Wk)		AHST 149,633 416

* Alaska-Hawaii Standard Time.

were combined to characterize the academic libraries and strata 4 and 5 were combined to characterize the public libraries. These combined strata are shown in the lower half of Figure 4A-1 along with the other networks. Thus, the lower half of Figure 4A-1 shows annual circulation data for each library category within each utility by time zone. Daily circulation was computed for each component cell in this table in the following way: annual circulation data for academic libraries were divided by 360 on the assumption that these libraries were open 7 days per week (closed on holidays), while public library circulation figures were divided by 313, on the assumption that they were open 6 days per week. (6)

(3) Daily academic circulation data from Figure 4A-1 were plotted on semi-log paper using the Lipetz card catalog traffic curve for academic libraries. The principal assumption was that circulation correlates one-to-one with card catalog use, at least for academic libraries. Figures 4A-2 thru 4A-4 show the estimated daily card catalog usage on an hour-by-hour basis for each utility. Also shown in each figure is the estimated cumulative card catalog traffic across all time zones.

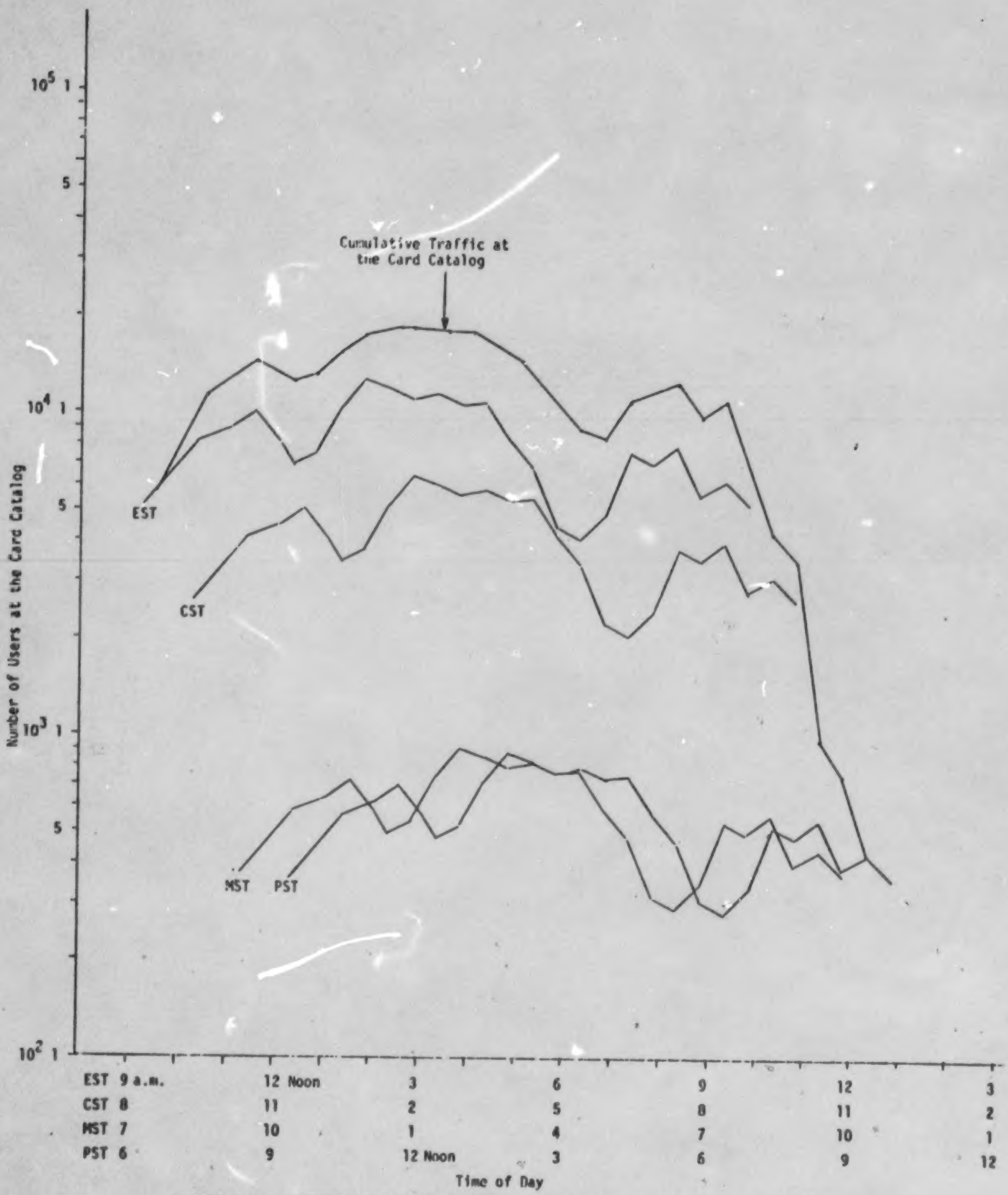


FIGURE 4A-2. ESTIMATED CARD CATALOG USE BY OCLC ACADEMIC LIBRARIES BY HOUR OF DAY

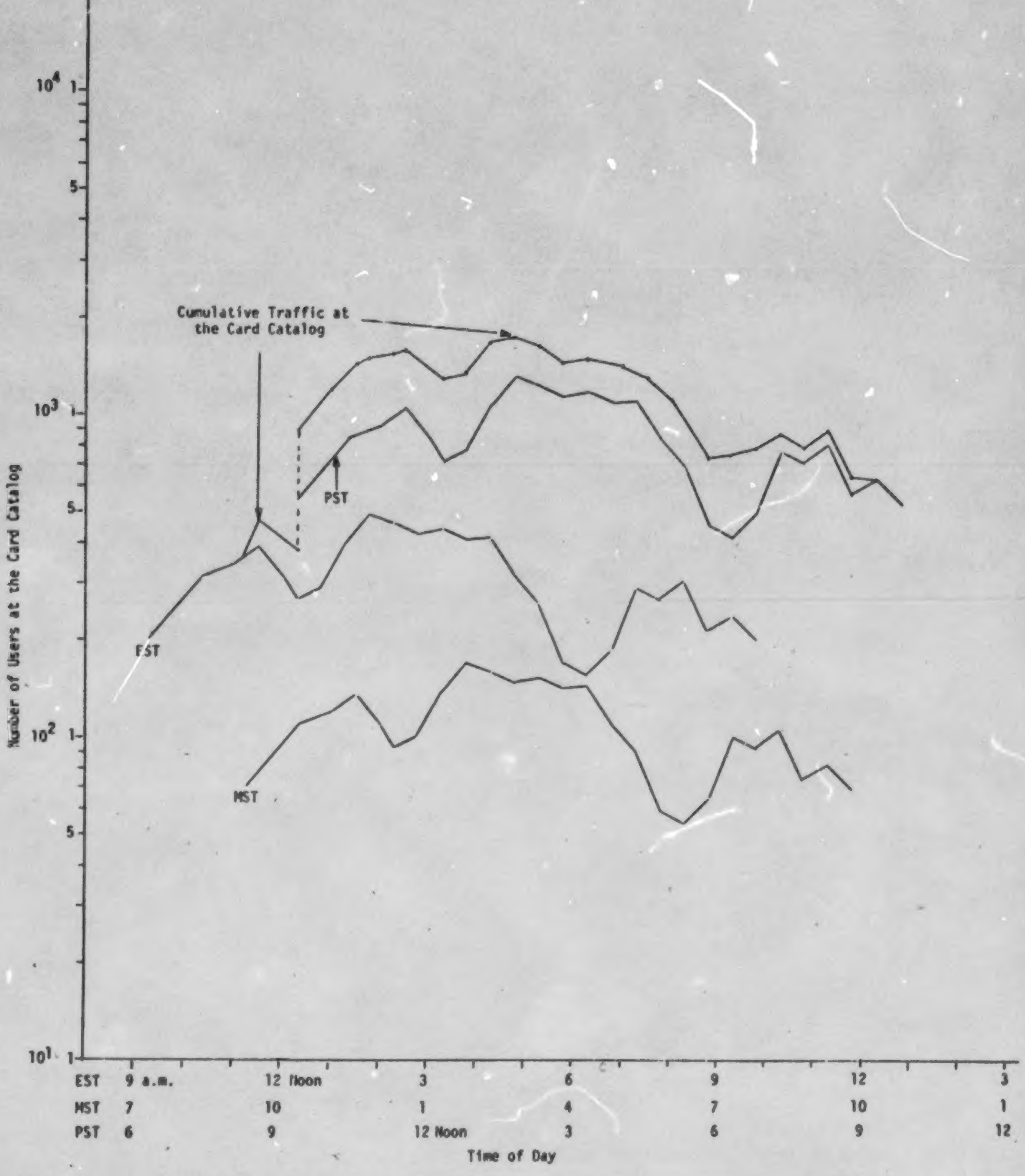


FIGURE 4A-3. ESTIMATED CARD CATALOG USE BY RLIN ACADEMIC LIBRARIES BY HOUR OF DAY

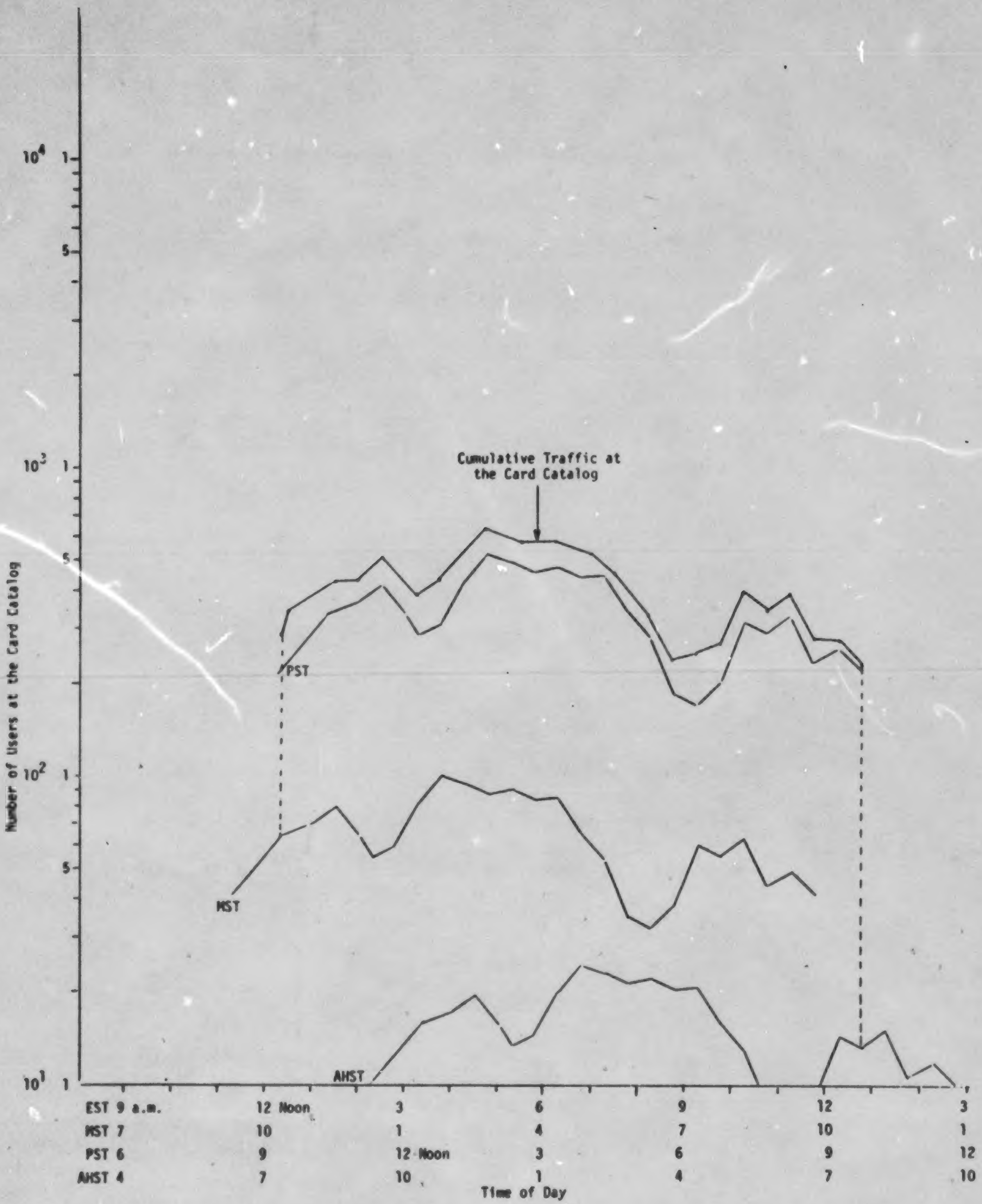


FIGURE 4A-4. ESTIMATED CARD CATALOG USE BY WLN ACADEMIC LIBRARIES BY HOUR OF DAY

(4) Daily public library circulation from Figure 4A-1 were plotted in a similar fashion, except that the Lipetz curve was replaced with the composite curve developed for public libraries from Butler's data. In addition, the public library circulation data shown in Figure 4A-1 were divided by 8 (that is, library card catalog use correlates on a one-to-eight basis, or one card catalog usage for every eight items circulated). Figures 4A-5 thru 4A-7 chart the findings.

(5) The cumulative traffic curves shown in Figures 4A-2 thru 4A-4 were combined in Figure 4A-3 to show possible combined catalog traffic in academic libraries across all utilities. Similarly, Figures 4A-5 thru 4A-7 were combined in Figure 4A-9 to show possible combined traffic in public libraries across all utilities.

(6) Figures 4A-8 and 4A-9 were combined in Figure 4A-10 to show the total estimated catalog usage in 1,227 utility-affiliated libraries.

Figure 4A-11 shows the estimated total card catalog use in the U.S. As can be seen, the data are not for comparable years. However, no

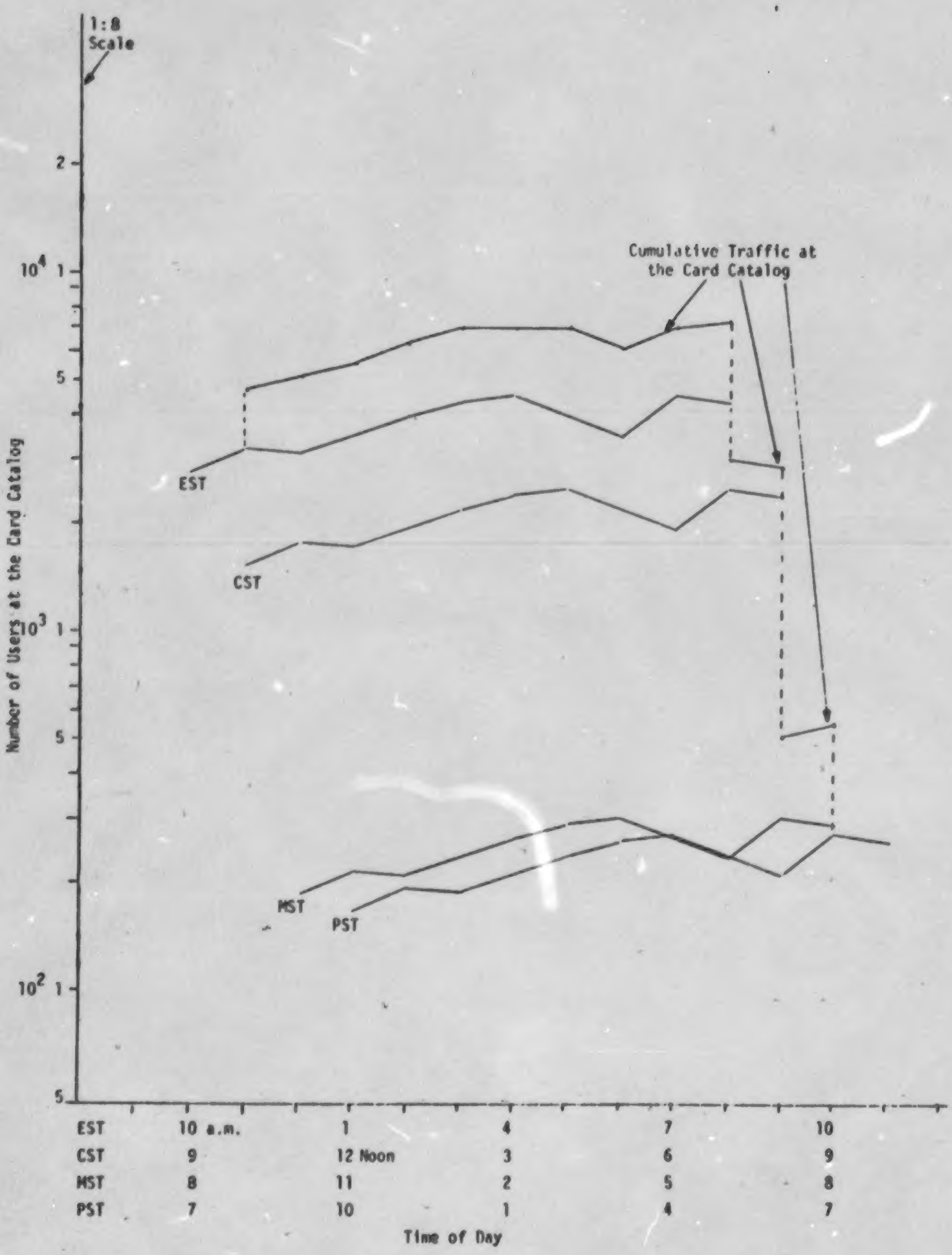


FIGURE 4A-5. ESTIMATED CARD CATALOG USE IN OCLC PUBLIC LIBRARIES BY TIME OF DAY

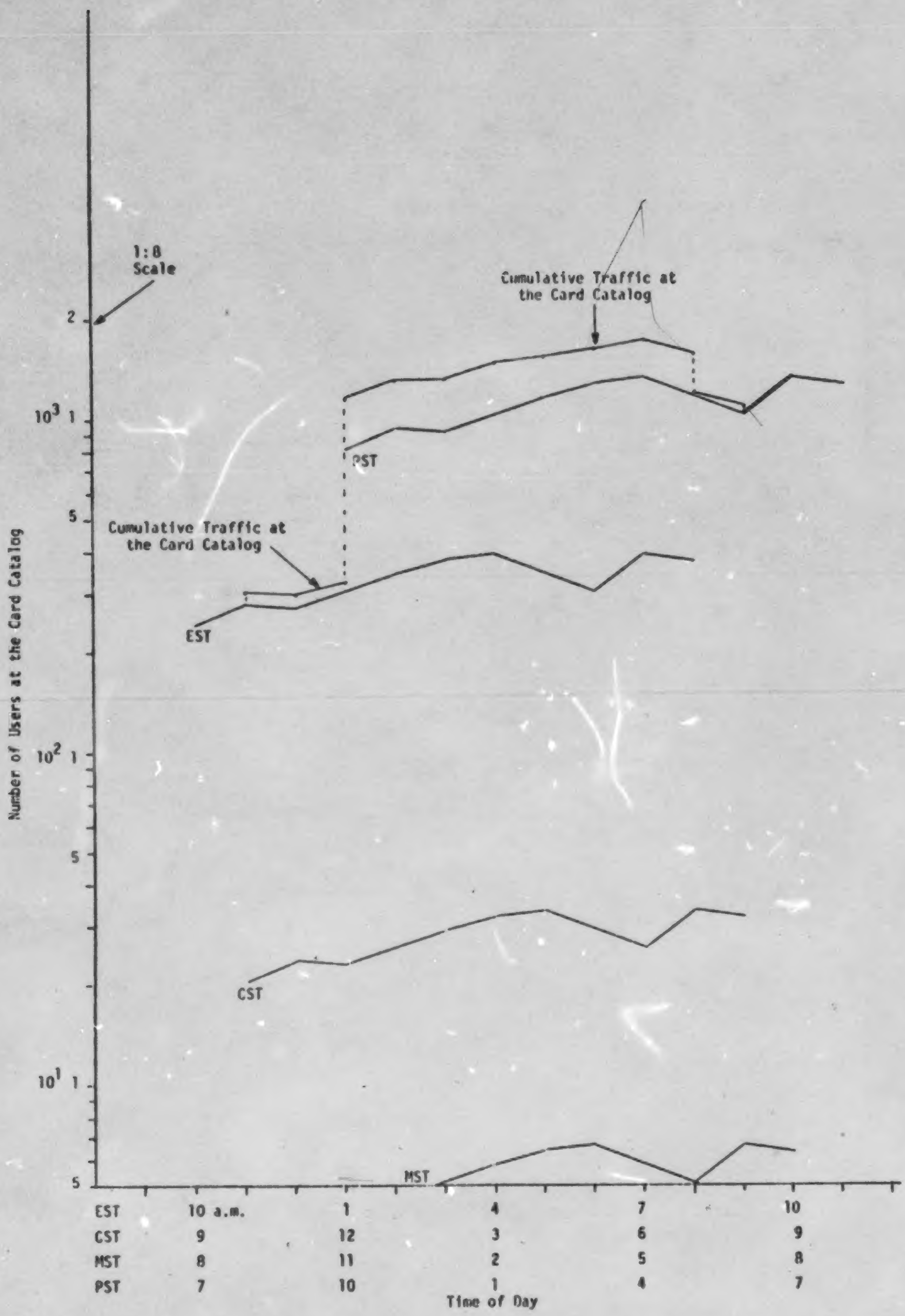


FIGURE 4A-6. ESTIMATED CARD CATALOG USE IN RLIN PUBLIC LIBRARIES BY TIME OF DAY

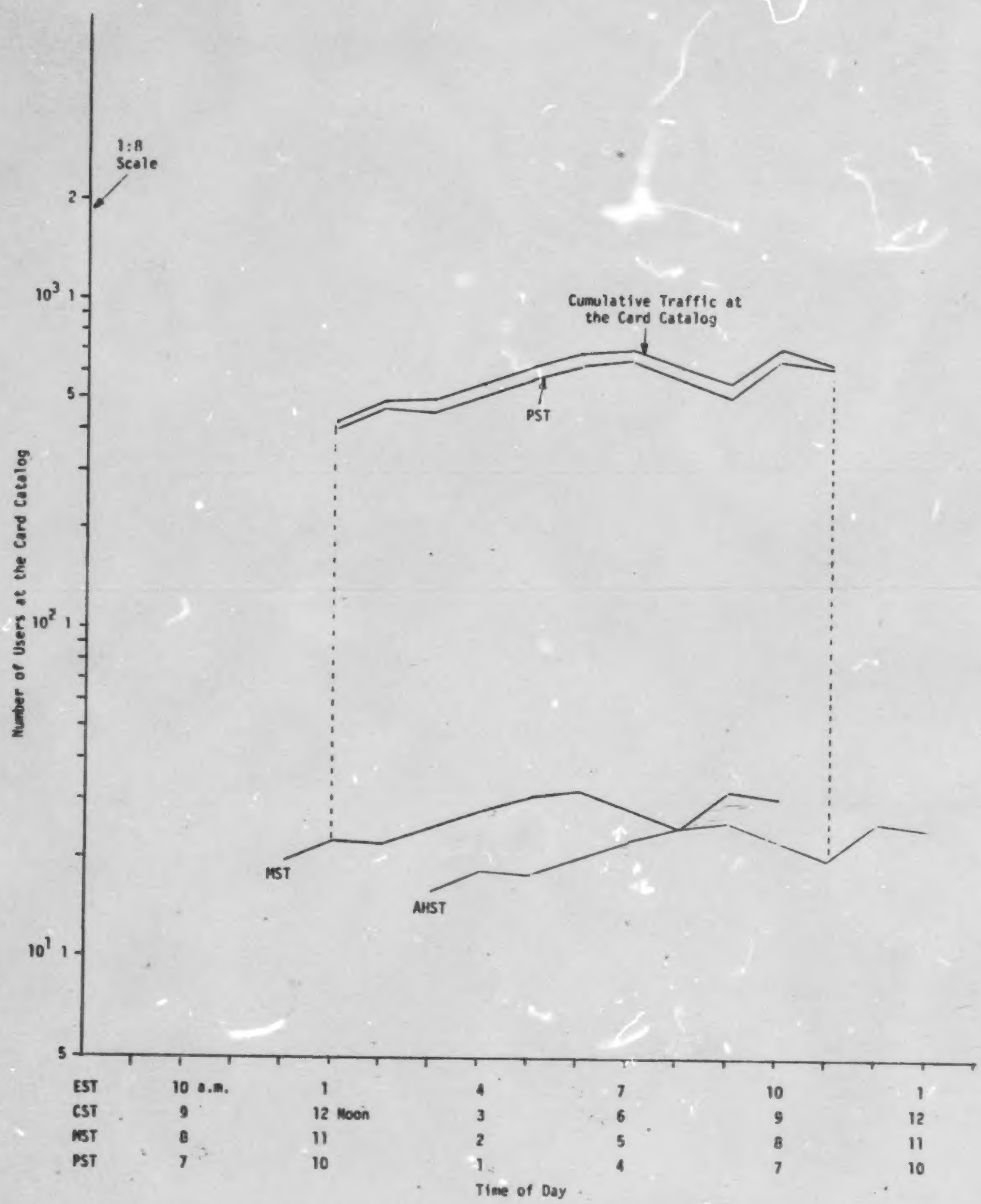


FIGURE 4A-7. ESTIMATED CARD CATALOG USE IN WLN PUBLIC LIBRARIES BY HOUR OF DAY

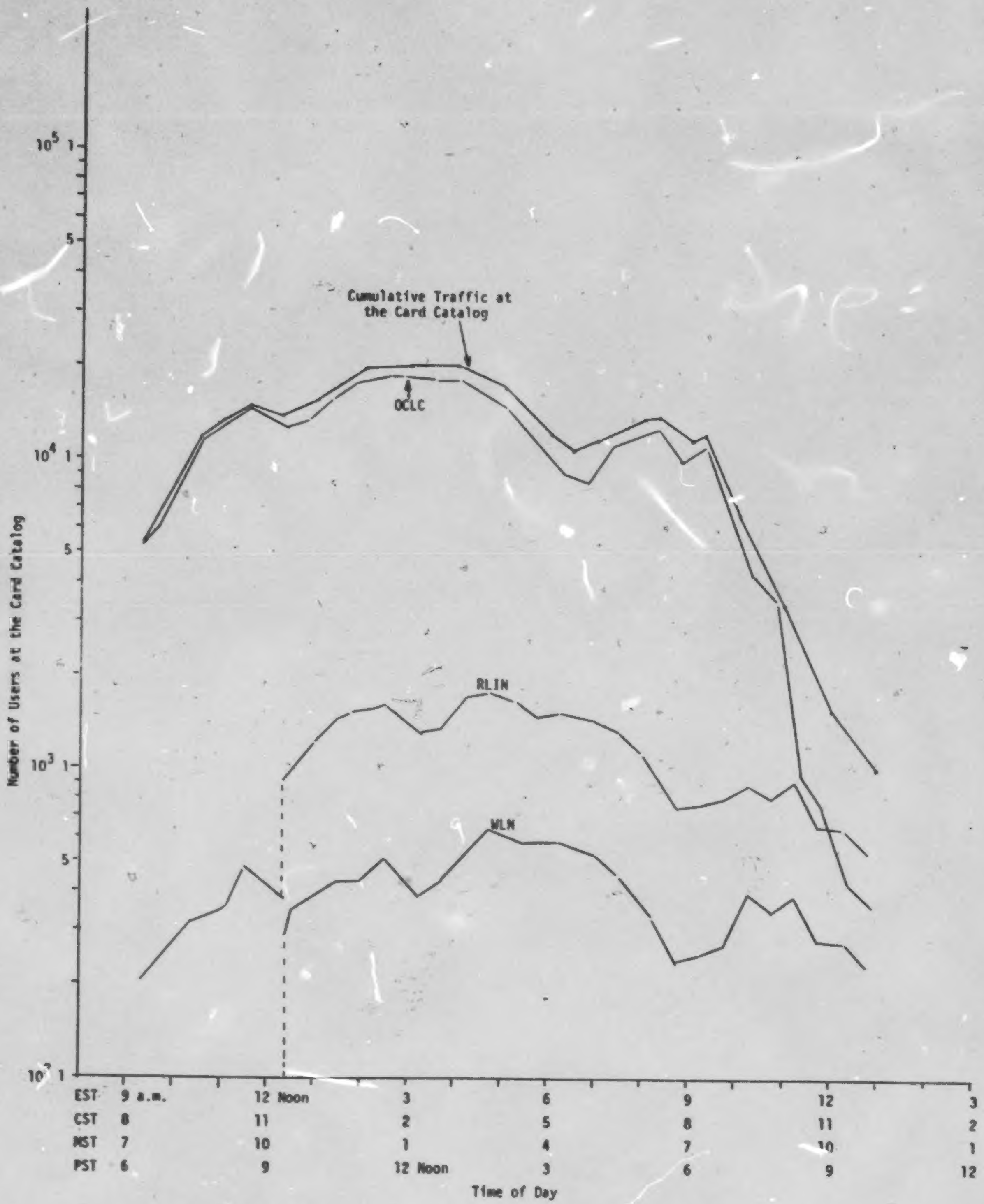


FIGURE 4A-8. ESTIMATED CARD CATALOG USE IN ACADEMIC LIBRARIES--ALL UTILITIES COMBINED

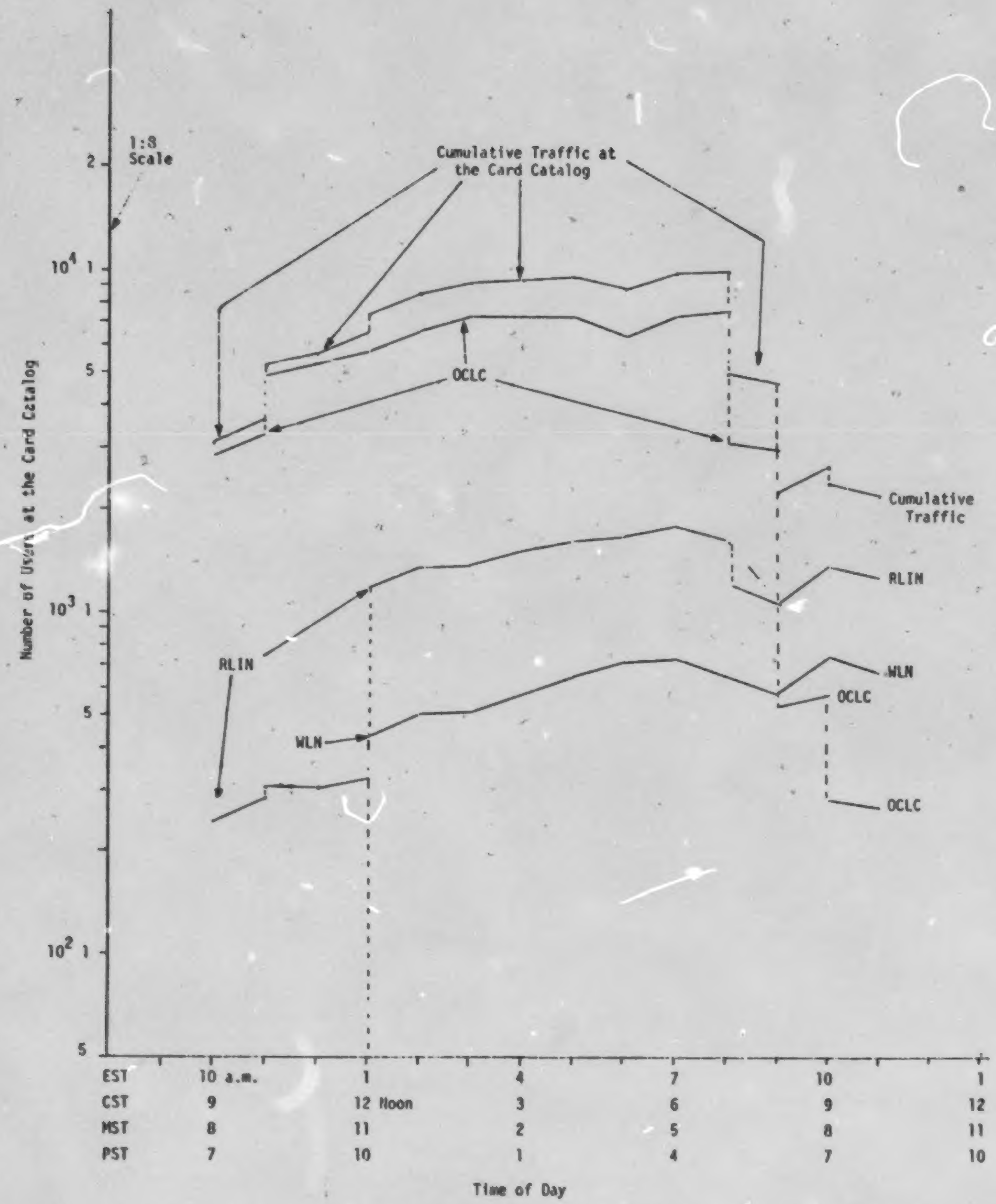


FIGURE 4A-9. ESTIMATED CARD CATALOG USE IN PUBLIC LIBRARIES--ALL UTILITIES COMBINED

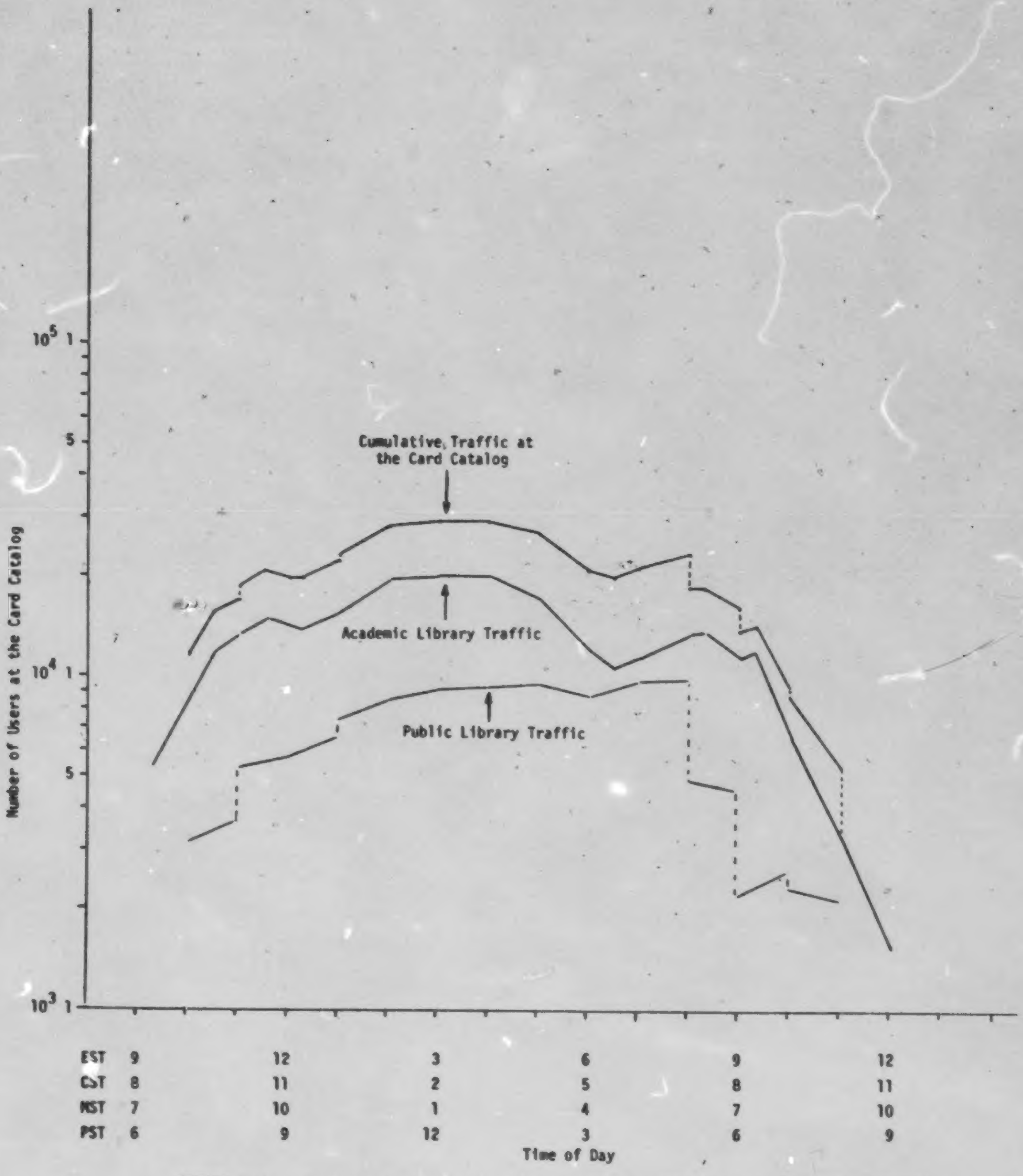


FIGURE 4A-10. ESTIMATED CARD CATALOG TRAFFIC IN MEMBER LIBRARIES--ALL UTILITIES COMBINED

FIGURE 4A-11. ESTIMATED DAILY CARD CATALOG USE IN THE U.S.

Type of Library	Annual Circulation	Source	Estimated Total Daily Card Catalog Use
(1) All Academic Libraries	198,719,411	NCES (1977)	552,000*
(2) All Public Libraries	892,852,268	LIBGIS I (1974)	357,000**
(3) All Libraries in U.S.	1,091,591,679		909,000

* 7-day week and 1:1 ratio of card catalog use with circulation.

** 6-day week and 1:8 ratio of card catalog use with circulation.

assumptions were made for growth or decline of circulation over time. An estimated 909,000 people are using card catalogs on an average day. This represents an estimated potential online reference search traffic due to usage of the card catalog--to be shared by the utilities in some fashion.

Some general observations about catalog traffic are noted for each utility based on the data presented in the figures.

OCLC

This utility is "driven" by eastern and central standard time libraries, both academic and public (Figures 4A-2 and 4A-5). Peak catalog traffic is estimated at about 18,000 users at 3 p.m. EST for academic libraries and about 7,500 users between 3 and 5 p.m. and again at 8 p.m. for public library users.

RLIN

This utility is greatly influenced by pacific and eastern standard time zone libraries, with public and academic libraries exhibiting nearly equal influence (Figures 4A-3 and 4A-6). Peak catalog traffic is estimated at about 1,800 users at 5 p.m. EST for academic libraries and about 1,800 users at 7 p.m. EST for public libraries.

WLN

This utility is dominated by PST libraries, with public libraries exhibiting greater influence than academic libraries (Figures 4A-4 and 4A-7). Peak catalog traffic is estimated at 650 users at about 5 p.m. EST for academic libraries and 800 users at about 10 p.m. EST for public libraries.

Linked Bibliographic Utilities

OCLC would be the dominant factor in linking the utilities for online card catalog use (Figures 4A-8 and 4A-9). When all libraries are considered, academic libraries contribute about twice as much traffic as public libraries during peak load (Figure 4A-10). Peak traffic is reached at about 3 p.m. EST, when about 30,000 users are at the card catalog.

Terminal Requirements

The objective of this subtask was to estimate the number of computer terminals needed to support general patron use of an online catalog assuming that card catalogs would be used very little. This estimate was limited to academic and public library members within the three bibliographic utilities. The task of estimating the number of terminals for a large network is great, and all estimates should be viewed as approximations. Although the analysis of this issue included precise results from the Theory of Queues, the raw data were less precise. The user of the suggested procedures should feel comfortable with the assumptions made about the nature of card catalog use, and thus the analytic framework. The analysis also suggests the type of data required in order to make full use of the estimation procedure.

In estimating the number of terminals needed for online reference searching, the first task is to disaggregate the data. A simultaneous consideration of the network-wide annual circulation of library materials or card catalog use clouds the problem considerably. Even though an aggregation of the library use data is necessary for bibliographic utility resource capacity assessment, it is not useful for an estimate of the number of terminals required. Different libraries have their busy periods at different times, and terminals cannot be shifted from one library to another easily, if at all. The result is that the number of terminals needed to satisfy the sum of all the worst cases is much greater than the number of terminals needed to satisfy the aggregate worst case for the network.

For these reasons, the analytical model described below was developed for use in a single library setting. The data collected for the catalog use study was disaggregated by size and type of library. For each of the five resulting strata, a "representative library" was constructed and the total catalog use divided evenly among the number of libraries in the stratum as if each one were equal to the representative case. The model was then run against the representative case with its representative traffic and the results multiplied by the number of libraries in the stratum to derive a total terminal requirement for all libraries. Below, the derivation of the model is discussed, the construction of the representative libraries is explained, and, finally, the findings for terminal requirements are presented.

The Model

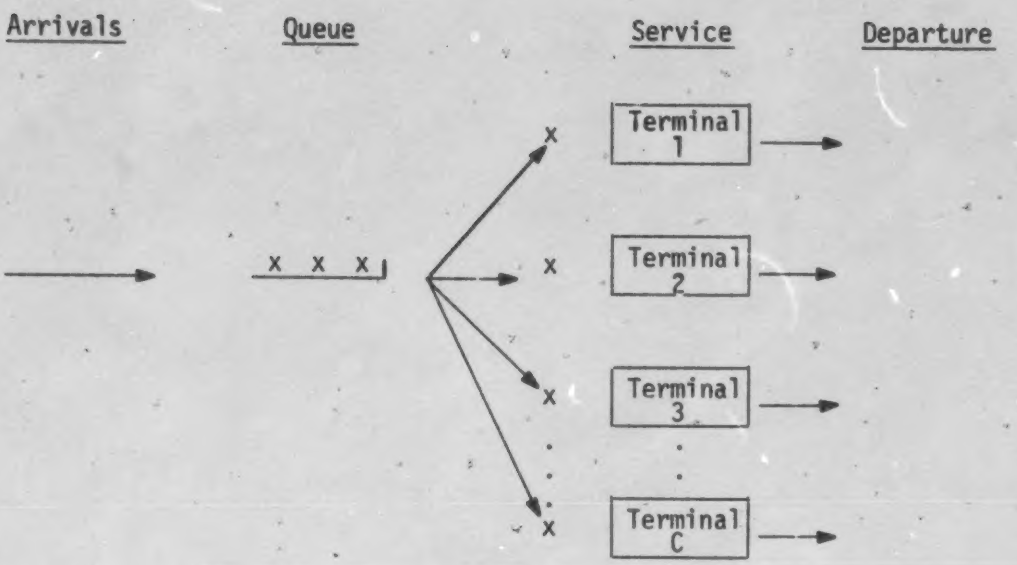
The model deemed to be most appropriate for estimating the number of terminals needed at a library for the worst case analysis (or busiest hour analysis) was the so-called multi-channel system model from the Theory of Queues, M/M/C (7). The M/M/C system is a queuing system that has C-parallel channels or servers. The assumptions are:

(1) Customer arrivals follow a Poisson distribution

(2) There are C identical servers (in this case terminals)

(3) The service time distribution for each terminal is exponential

(4) Arriving patrons use the first available terminal and, if all terminals are in use, a single unlimited queue forms.


Arrivals
Queue
Service
Departure
x x x
x
x
x
x
Terminal 1
Terminal 2
Terminal 3
Terminal C

The assumption of Poisson arrivals seems well-justified from an intuitive sense and from available data (8). The Poisson process has the following characteristics:

- Completely random arrivals
- Customers completely independent of one another
- The number of arriving customers is proportional to elapsed time.

All studies reviewed agreed in principle that the arrival process at library catalogs is indeed Poisson. Although the arrival process was modeled as Poisson, the actual arrival rate varied with time of day. Libraries have more arrivals during some hours of the day than other hours.

For estimating purposes, the worst-case or busiest-hour case was considered; thus the suggested rate of arrivals to be used is the arrival rate during the peak hour of the day. Let λ be the arrival rate at the busiest time of the day for a library.

The second assumption, that of C identical servers, is easily justified. If all terminals within a given library are similar and have equal access to the computer, then the assumption is valid.

The third assumption, exponential service time distribution, also seems well justified (9). Work by Ronald W. and Jo Ellen Force (8), graphically demonstrates that the exponential distribution is a good choice. An assumption that the average service rate does not vary was made and seems quite reasonable. Note that the assumption that the average service rate does not vary presupposes that the computer network configuration is adequate to meet the demands of the network; i.e., response time is not noticeably degraded by the number of concurrent users.

The fourth assumption concerns queue discipline. If a single queue forms behind the C terminals, this condition is met. Note that the single queue is in fact the best design since wait-times for individual customers have minimum variance in this set-up. If C queues formed, one behind each terminal, the average wait-time for a customer would be the same as in the single queue set-up, but the variance of customer wait-time would be larger. The single queue model, therefore, seems to be most reasonable.

It should be noted that this analysis made no assumptions about terminal placement within a single library--that is, the number of clusters (or banks) of terminals located throughout the library building. Consequently, our estimates may be somewhat conservative.

The model will be presented in terms of a few analytical formulae. Variables to be defined are:

(1) C = Number of terminals at a library

(2) λ = Arrivals during the busiest part of the day

(3) μ = Service rate of a single terminal (terminal session time per patron).

The model will be described by a formula for the probability of m in the system as well as expected time in the system and expected number in the system.

Probability (of m customers in the system) = p_m $\quad m = 0,1,2,\ldots$

①

$$p_m = \begin{cases} \dfrac{(\lambda/\mu)^m}{m!} \cdot p_0 & ; \quad 0 \le m \le C \\[2ex] \dfrac{(\lambda/\mu)^m}{C!\,C^{m-C}} \cdot p_0 & ; \quad m > C \end{cases}$$

② where $$p_o = \left[\sum_{m=o}^{c-1} (\lambda/\mu)^m / m! + \frac{(\lambda/\mu)^c}{C!\,(1-\frac{\lambda}{c\mu})} \right]^{-1}$$

p_o = probability (of no one in the system)

L_q = expected number of customers in the queue

③ $$L_q = \frac{(\lambda/\mu)^c \left(\lambda/(c\mu)\right)}{C!\,\left(1-\lambda/(c\mu)\right)^2} \cdot p_o$$

W_q = expected amount of time in the queue for a customer

④ $$W_q = L_q/\lambda$$

The Cumulative Waiting Time Distribution

⑤ $$W_q\,(t) = \left\{ \begin{array}{lll} 1-\left[\frac{C(\lambda/\mu)^c}{C!(C-\lambda/\mu)} \cdot p_o\right] & ; & t = o \\ \frac{(\lambda/\mu)^c \left(1-\exp\left(-t\ \mu c-\lambda\right)\right)}{(c-1)!\ (c-\lambda/\mu)}\Big) . p_o + W_q(o) & & t > o \end{array} \right\}$$

Suggested Use of Model

The Battelle data do not include any arrival rate data; however, card catalog use data are included. It will be assumed that the hourly fluctuations in card catalog use are similar in all libraries--that is, the peak arrival rate is approximately 1.59 times the average hourly arrival rate. In the Force paper (9), the daily arrival rate was 160.6 in 14 hours, for an average hourly arrival rate of 11.47. The maximum hourly arrival rate was 18.2 per hour. So the ratio of maximum hourly arrival rate to the average daily arrival rate was 18.2/11.47 = 1.587. Data from the Columbia University Library demonstrated the same behavior.

Since the card catalog is similar to a queuing system with an unlimited number of servers, the ratio of maximum observed customers to average observed customers should be very similar to the ratio of maximum arrival rate to average hourly arrival rate. In summary, the proposed method for estimating arrival rate for worst case analysis is as follows:

Academic Libraries	Public Libraries
(1) Annual Circulation = Y	Annual Circulation = Y
(2) Daily Circulation: $D = Y \div 360$	$D = Y \div 313$
(3) Average Hourly Circulation: $H = D \div 14$	$H = D \div X$*
(4) Busy Arrival Rate: $\lambda = H \times 1.59$	$\lambda = H \times 1.19$

* X = 12 for large public libraries; 10 for small public libraries.

Service Time Rates

Studies at Northwestern University and Ohio State University indicate that card catalog service times average about 7 minutes. From the Knox and Miller (9) data, an average of 6.6 minutes was reported (although their system is presently online), while OSU studies suggested rates of between 7.06 and 7.51 minutes. A range of service times was selected for the model in order to bracket the average service time at a card catalog. This was done to accommodate possible variances in service time as a result of an online catalog.

Since the circulation data used in this analysis were not collected with a queuing analysis in mind, some problems exist in using the figures for this purpose. For instance, if the OSU yearly circulation data are used to compute a worst case arrival rate, a poor figure will result. This is primarily because the OSU library system consists of more than 20 individual libraries, so the variance in the data is hidden. Smaller schools with just one library will yield more accurate results. The larger 'libraries' usually include branches, or departmental libraries, in order to distribute the service load. Thus, as the estimated library size (in terms of circulation or use of the card catalog) increases, the accuracy of this model decreases. In the absence of individual data for each branch or departmental library, it was necessary to make some general assumptions. In the OSU case, for example, one might assume that 20 libraries exist and that the main library captures 30 percent of the catalog traffic. Now the worst case arrival rate for OSU-Main is computed by allocating the total estimated catalog traffic to the main library and to the branches, then applying the worst case arrival rate formulas described above to the resulting traffic figures. The computation for OSU-main then becomes $Y_M = (Y \times .30)$. Thus, a worst case analysis applied to the OSU library catalog traffic figure and the resulting value for λ may be apportioned at the rate of 30 percent to main, 70 percent to the branches.

In general, the larger the library, the smaller the percentage of its business done in its main branch.

Model Configuration

Figure 4A-12 shows the representative configurations assumed for the five strata of libraries in the bibliographic networks. The following estimates were prepared for each stratum: the number of libraries represented (total and by network); the total annual circulation; the average annual circulation per library (line 2 divided by line 1); the average daily circulation by library (assuming 360 days per year for academic and 313 days per year for public libraries); the average hourly circulation (assuming 14 hours per day for academic and 10-12 hours per day for public libraries); and the worst case hourly circulation for academic libraries. In the case of public libraries, the average annual circulation must be further divided by 8 to achieve the average annual catalog use.

The hourly traffic figures shown were divided among the assumed constituent parts of the "representative library" as follows:

FIGURE 4A-12. CIRCULATION (CARD CATALOG USE) BY TYPE OF LIBRARY BASED ON UTILITY AFFILIATION

	OCLC	RLIN	WLN	TOTAL	ASSUMED LIBRARY CONFIGURATIONS
Research Libraries					
Number of libraries	68	14	2	84	Main library captures 40% of traffic; Five large branches capture 30% of traffic; Fifteen small branches capture 30% of traffic
Annual circulation	47,970,000	12,479,000	1,627,000	62,076,000	
Annual circulation/library	705,000	891,000	814,000	739,000	
Daily circulation/library (360 d)				2,053	
Ave. hourly circulation (14 hr)				147	
Worst case hourly circulation*				183+	
Large Academic Libraries					
Number of libraries	526	8	13	547	Main library captures 70% of traffic; Five large branches capture 20% of traffic; Five small branches capture 10% of traffic
Annual circulation	60,129,000	4,604,000	2,011,000	66,744,000	
Annual circulation/library	114,000	576,000	155,000	122,000	
Daily circulation/library (360 d)				340	
Ave. hourly circulation (14 hr)				24	
Worst case hourly circulation				30+	
Small Academic Libraries					
Number of libraries	369	3	20	392	Main library captures 100% of traffic
Annual circulation	17,158,000	170,000	957,000	18,285,000	
Annual circulation/library	46,000	57,000	48,000	47,000	
Daily circulation/library (360 d)				130	
Ave. hourly circulation (14 hr)				9	
Worst case hourly circulation				11+	
Large Public Libraries					
Number of libraries	82	9	13	104	Main library captures 15% of traffic; Twenty branches capture 85% of traffic
Annual circulation	150,211,000	38,762,000	14,238,000	203,211,000	
1/8 annual circulation	18,776,000	4,845,000	1,780,000	25,401,000	
1/8 annual circulation/library	229,000	538,000	137,000	244,000	
Daily circulation/library (313 d)				780	
Ave. hourly circulation (12 hr)				65+	
Small Public Libraries					
Number of libraries	92	3	5	100	Main library captures 100% of traffic
Annual circulation	31,816,000	1,309,000	1,730,000	34,855,000	
1/8 annual circulation	3,977,000	164,000	216,000	4,357,000	
1/8 annual circulation/library	43,000	55,000	43,000	44,000	
Daily circulation/library (313 d)				141	
Ave. hourly circulation (10 hr)				14+	

* Worst case hourly circulation is defined as hourly circulation based on weekday card catalog use in non-summer months. Average hourly circulation was multiplied by 124.7% to arrive at worst cases for academic libraries. This conversion was not necessary for public libraries.

+ Represents per-hour card catalog traffic per library system, to be allocated according to the assumed library configuration and supplied as lambda (λ) to the queuing model.

- Research Libraries--40 percent of the catalog traffic was allocated to the main library, an inference from figures given by Lipetz. Five large branches were postulated as accounting for 30 percent of the total traffic, 15 smaller branches account for 30 percent.

- Large Academic--It was assumed that one main library captures 70 percent of total traffic, five large branches capture 20 percent, five small branches take 10 percent. Since this stratum includes all academic libraries with more than 100,000 volumes (except for ARL members), there will be wide variability within the stratum. Any single library may be quite different from the chosen representative.

- Small Academic--Assumed to be a single main library only.

- Large Public--One main library is assumed to capture 15 percent of all circulation, with 20 branches capturing 85 percent.

- Small Public--Assumed to be one library location only.

Determination of Terminal Requirements

With the average hourly traffic, and a plausible traffic configuration for each type of library, the queuing model was used to estimate the number of computer terminals for each library node in each configuration. To simplify the estimation, it was decided to solve for C (the number of terminals at a library) by supplying only λ (arrivals during busiest part of the day) and μ (the service rate of a single terminal in the library).

To allow for variability in certain key factors of the analysis, ranges of values were explored. These factors included wait time (N) in a queue and service rate (μ) at the terminal. The values chosen for wait time in a queue were 0 (no waiting), 2, 4, and 6 minutes; queuing solutions were derived for each case. However, Battelle's analysis focused on a queue wait time of 2 minutes or less on the intuitive assumption that in a library setting a significant amount of balking (refusal to enter a long line) or reneging (leaving the queue after waiting for a certain amount of time) would occur beyond a 2-minute wait for a terminal.

A range of service rates was also chosen for the model, due to insufficient data concerning a patron's total session time at a public access terminal. Furthermore, the results of past studies of card catalog search times varied considerably, from 2 to 7.5 minutes per visit. Battelle's analysis of unpublished data from the Ohio State University main library showed the average card catalog search (including staff usage) to be 7.5 minutes for 6,595 searches. Finally, it was concluded that even if better knowledge existed about card catalog search times, there would be no way of estimating the conversion of card catalog searches into comparable online catalog searches. The service rates at public access terminals selected for the queuing model were 2, 4, 8, and 12 minutes. The Battelle team considered 4 minutes to be a reasonable service rate (terminal session time per patron) for purposes of this analysis.

The queuing equations described above were programmed for the simplified model, and solved for C (number of terminals) using a computer. Results are shown in Figure 4A-13. Lambda (λ) was computed from Figure 4A-12, while Mu (μ) and queue wait time (N) were assigned a range of values as previously described. The model estimated C (number of terminals) and the probabilities of waiting in a queue N minutes or less. The only solutions shown in Figure 4A-13 are those where the probability of waiting in a queue N minutes or less was 90 percent or greater.

Figures 4A-14 thru 4A-17 show the estimated terminal requirements for individual utilities and for all utilities combined. These estimates were prepared by simple summation of the results from Figure 4A-13. Thus, to arrive at the number of terminals needed to provide online catalog searching at research libraries in OCLC, assuming a 4-minute service time and a 90 percent or greater probability that the wait time in a queue will be 2 minutes or less, the summation consisted of 11 terminals for each of 68 main libraries (see Figure 4A-13) plus 3 terminals for each of 340 large branches (5 large branches at each of 68 libraries) plus 2 terminals for each of 1,020 small branches (15 small branches at each of 68 libraries). The result was 3,810 terminals for 68 research libraries in the OCLC network.

FIGURE 4A-13. NUMBER OF COMPUTER TERMINALS REQUIRED PER LIBRARY TO HANDLE ONLINE CATALOG TRAFFIC AT BUSIEST HOUR(S) OF THE DAY

RESEARCH LIBRARY SYSTEM	Peak Traffic* (λ)	Service Rate (in min.) (μ)	Probability** of Waiting in Queue N Minutes or Less				# Terminals per Library (C)
			Range of N's				
			no waiting	≤ 2 min.	≤ 4 min.	≤ 6 min.	
Main Library	116	2			93.7	97.8	5
	116	②		96.6			⑥
	116	2	94.6				8
	116	4			95.5	98.5	10
	116	④		95.3			⑪
	116	4	93.2				13
	116	8				91.2	18
	116	8			93.5		19
	116	⑧		92.2			⑳
	116	8	90.4				22
	116	12				92.8	27
	116	12			93.3		28
	116	⑫		91.5			㉙
	116	12	93.1				32

○ Queuing solutions selected by Battelle for subsequent estimations.

* Estimated number of arrivals per hour at online catalog at busiest time of day. Peak load for academic libraries is 1.59 times the worst case hourly traffic, apportioned according to the assumed representative library configuration shown in Figure 4A-12. Peak load for public libraries is 1.19 times the average hourly traffic.

** Only probabilities of 90% or greater are shown in the table, and then only the solution associated with the fewest number of terminals.

FIGURE 4A-13
(Continued)

RESEARCH LIBRARY SYSTEM	Peak Traffic* (λ)	Service Rate (in min.) (μ)	Probability** of Waiting in Queue N Minutes or Less				# Terminals per Library (C)
			Range of N's				
			No waiting	≤ 2 min.	≤ 4 min.	≤ 6 min.	
Large Branch	14	2				90.6	1
	14	②	91.2	[✓ 98.1]	99.6		②
	14	4				94.0	2
	14	④	92.3	[✓ 97.3]	99.0		③
	14	⑧		[✓ 91.0]	95.1	97.1	④
	14	8	95.4				5
	14	12			90.9	93.7	5
	14	⑫	92.5	[✓ 95.6]			⑥
Small Branch	5	②		[✓ 92.8]	96.9	98.6	①
	5	2	98.7				2
	5	④	95.2	[✓ 97.9]	99.1	99.6	②
	5	8			91.4	93.9	2
	5	⑧	96.8	[✓ 98.2]			③
	5	⑫	90.9	[✓ 93.5]	95.3	96.7	③
LARGE ACADEMIC LIBRARY SYSTEM							
Main Library	33	2			93.6	97.4	2
	33	②		[✓ 98.3]			③
	33	2	97.2				4
	33	④		[✓ 90.8]	96.3	98.5	④
	33	4	91.6				5
	33	8			94.6	97.2	7
	33	⑧	90.7	[✓ 96.2]			⑧
	33	12				91.0	9
	33	⑫		[✓ 90.6]	94.7		⑩
	33	12	91.3				11
Large Branch	5	②		[✓ 92.8]	96.9	98.6	①
	5	2	98.7				2
	5	④	95.2	[✓ 97.9]	99.1	99.6	②
	5	8			91.4	93.9	2
	5	⑧	96.8	[✓ 98.2]			③
	5	⑫	90.9	[✓ 93.5]	95.3	96.7	③
Small Branch	1	②	--	--	--		①
	1	④	--	--	--		①
	1	⑧	--	--	--		①
	1	⑫	--	--	--		①

FIGURE 4A-13
(Continued)

SMALL ACADEMIC LIBRARY SYSTEM	Peak Traffic* (λ)	Service Rate (in min.) (μ)	Probability** of Waiting in Queue N Minutes or Less				# Terminals per Library (C)
			Range of N's				
			No waiting	≤ 2 min.	≤ 4 min.	≤ 6 min.	
Main Branch	17	②		97.2	99.3	99.8	②
	17	2	97.9				3
	17	④		95.2	98.1	99.3	③
	17	4	96.9				4
	17	8				93.3	4
	17	⑧	90.7	95.3	97.6		⑤
	17	12			93.3	95.7	6
	17	12	93.3	96.3			⑦
No Branches							
LARGE PUBLIC LIBRARY SYSTEM							
Central Library	12	2				93.4	1
	12	②	93.3	98.7	99.7		②
	12	4			93.1	96.2	2
	12	④	94.8	98.3			③
	12	8				90.4	3
	12	⑧	90.9	95.0	97.3		④
	12	12		92.6	95.2	96.9	⑤
	12	⑫	96.0				6
Branch Library	3	②	90.0	95.9	98.4	99.3	①
	3	4			91.0	94.0	1
	3	④	98.2	99.3			②
	3	⑧	93.3	95.5	97.0	98.0	②
	3	12			91.3	93.1	2
	3	⑫	97.5	98.4			③
SMALL PUBLIC LIBRARY SYSTEM							
Central Library	17	②		97.2	99.3	99.8	②
	17	2	97.9				3
	17	④		95.2	98.1	99.3	③
	17	4	96.9				4
	17	8				93.3	4
	17	⑧	90.7	95.3	97.6		⑤
	17	12			93.3	95.7	6
	17	⑫	93.3	96.3			⑦
No Branches							

FIGURE 4A-14. ESTIMATED NUMBER OF COMPUTER TERMINALS NEEDED TO SUPPORT ONLINE CATALOGING SEARCHING

Academic and Public Libraries
in OCLC Network

Size and Type of Library	NUMBER OF TERMINALS* NEEDED AT VARIOUS SERVICE RATES**			
	2 minutes	4 minutes	8 minutes	12 minutes
68 Research Libraries	2,110+	3,810	5,780	7,610
526 Large Academic Libraries	5,260	6,840	9,990	11,050
369 Small Academic Libraries	740	1,110	1,850	2,580
82 Large Public Libraries	1,800	3,530	3,610	5,330
92 Small Public Libraries	180	280	460	640
1,137 Total Libraries	10,090	15,570	21,690	27,210

*Based on a 90% (or greater) probability that waiting time (N) will be 2 minutes or less.

**Service rate (μ) is defined as terminal session time per patron.

+Figures are rounded off to nearest 10.

FIGURE 4A-15. ESTIMATED NUMBER OF COMPUTER TERMINALS NEEDED TO SUPPORT ONLINE CATALOG SEARCHING

Academic and Public Libraries
in RLIN Network

Size and Type of Library	NUMBER OF TERMINALS* NEEDED AT VARIOUS SERVICE RATES**			
	2 minutes	4 minutes	8 minutes	12 minutes
14 Research Libraries	435+	785	1,190	1,455
8 Large Academic Libraries	80	105	150	170
3 Small Academic Libraries	5	10	15	20
9 Large Public Libraries	200	385	395	580
3 Small Public Libraries	5	10	15	20
37 Total Libraries	725	1,295	1,765	2,245

*Based on a 90% (or greater) probability that waiting time (N) will be 2 minutes or less.

**Service rate (μ) is defined as terminal session time per patron.

+Figures are rounded off to nearest 10.

FIGURE 4A-16. ESTIMATED NUMBER OF COMPUTER TERMINALS NEEDED TO SUPPORT ONLINE CATALOG SEARCHING

Academic and Public Libraries
in WLN Network

Size and Type of Library	NUMBER OF TERMINALS* NEEDED AT VARIOUS SERVICE RATES** 2 minutes	4 minutes	8 minutes	12 minutes
2 Research Libraries	60+	110	170	210
13 Large Academic Libraries	130	170	245	275
20 Small Academic Libraries	40	60	100	140
13 Large Public Libraries	285	560	570	845
5 Small Public Libraries	10	15	25	35
53 Total Libraries	525	915	1,110	1,505

*Based on a 90% (or greater) probability that waiting time (N) will be 2 minutes or less.

**Service rate (μ) is defined as terminal session time per patron.

+Figures are rounded off to nearest 10.

FIGURE 4A-17. ESTIMATED NUMBER OF COMPUTER TERMINALS NEEDED TO SUPPORT ONLINE CATALOG SEARCHING

Academic and Public Libraries
in all Networks Combined

Size and Type of Library	NUMBER OF TERMINALS* NEEDED AT VARIOUS SERVICE RATES**			
	2 minutes	4 minutes	8 minutes	12 minutes
84 Research Libraries	2,605+	4,705	7,140	9,275
547 Large Academic Libraries	5,470	7,115	10,385	11,495
392 Small Academic Libraries	785	1,180	1,965	2,740
104 Large Public Libraries	2,285	4,475	4,575	6,755
100 Small Public Libraries	195	305	500	695
1,227 Total Libraries	11,340	17,780	24,565	30,960

[*] Based on a 90% (or greater) probability that waiting time (N) will be 2 minutes or less.

[**] Service rate (μ) is defined as terminal session time per patron.

[+] Figures are rounded off to nearest 10.

Computer Resources

The third task was to estimate the amount of processing required at each utility in order to support the catalog usage at member libraries. The processing will be expressed in terms of CPU seconds required to complete the projected number of searches.

Each utility was treated separately since each has a different processing configuration and a different search load. Figures 4A-18, 4A-19, and 4A-20 show the resulting CPU requirements for this application for OCLC, RLIN, and WLN, respectively.

OCLC

To construct Figure 4A-18, two sets of calculations were necessary--one for academic and one for public libraries. The process will be described fully for the derivation of academic library figures. Subsequent calculations can be explained as variations of this original process.

In accordance with the academic library catalog use curve and the direct correlation observed between circulation and catalog use, circulation data were used to plot hour-by-hour traffic at all member libraries. These cumulative totals are given by hour under Academic Uses in Figure 4A-18.

FIGURE 4A-18. COMPUTER RESOURCES FOR ONLINE CATALOG--OCLC CPU SECONDS ROUNDED TO NEAREST 10 SECONDS

Time (EST)	Academic Uses (x .73)	CPU Sec.	Public Uses (x .51)	CPU Sec.
9 a.m.	5,800	4,230		
10 a.m.	11,300	8,250	2,800	1,430
11 a.m.	14,620	10,670	4,800	2,450
12 p.m.	12,740	9,300	5,190	2,650
1 p.m.	15,450	11,280	5,740	2,930
2 p.m.	18,520	13,520	6,520	3,330
3 p.m.	18,120	13,230	7,190	3,670
4 p.m.	17,990	13,130	7,190	3,670
5 p.m.	14,660	10,700	7,250	3,700
6 p.m.	8,960	6,540	6,430	3,280
7 p.m.	11,030	8,050	7,250	3,700
8 p.m.	12,460	9,100	7,580	3,870
9 p.m.	11,020	8,040	2,900	1,480
10 p.m.	4,280	3,120	560	290
11 p.m.	980	720	260	130
12 a.m.	430	310	None	None
Total CPU Seconds		130,190		36,580
Total CPU Minutes		2,170		610
Total CPU Seconds (Worst Case)		162,350		36,580
Total CPU Minutes (Worst Case)		2,700		610

FIGURE 4A-19. COMPUTER RESOURCES FOR ONLINE CATALOG--RLIN
CPU SECONDS ROUNDED TO NEAREST SECOND

Time (EST)	Academic Uses	(x .075)	CPU Sec.	Public Uses	(x .09)	CPU Sec.
10 a.m.	260		20	250		23
11 a.m.	340		26	310		28
12 p.m.	420		32	320		29
1 p.m.	1,200		90	1,160		104
2 p.m.	1,550		116	1,340		121
3 p.m.	1,310		98	1,350		122
4 p.m.	1,680		126	1,480		133
5 p.m.	1,660		125	1,570		141
6 p.m.	1,505		113	1,630		147
7 p.m.	1,430		107	1,770		159
8 p.m.	1,125		84	1,600		144
9 p.m.	760		57	1,040		94
10 p.m.	885		66	1,320		119
11 p.m.	900		68	1,250		113
12 a.m.	640		48	None		None
Total CPU Seconds			1,176			1,477
Total CPU Minutes			20			25
Total CPU Seconds (Worst Case*)			1,466			1,477
Total CPU Minutes (Worst Case)			25			25

* Worst case = total CPU seconds X 124.7%.

FIGURE 4A-20. COMPUTER RESOURCES FOR ONLINE CATALOG--WLN
CPU SECONDS ROUNDED TO NEAREST SECOND

Time (EST)	Academic Uses	(x .22) CPU Sec.	Public Uses	(x .27) CPU Sec.
1 p.m.	345	76	430	116
2 p.m.	435	96	490	132
3 p.m.	420	92	500	135
4 p.m.	500	110	560	151
5 p.m.	635	140	640	173
6 p.m.	585	129	700	189
7 p.m.	535	118	720	194
8 p.m.	335	74	640	173
9 p.m.	250	55	570	154
10 p.m.	400	88	720	194
11 p.m.	390	86	650	176
12 a.m.	280	62	25	7
Total CPU Seconds		1,126		1,794
Total CPU Minutes		19		30
Total CPU Seconds (Worst Case)		1,404		1,794
Total CPU Minutes (Worst Case)		24		30

These uses were then multiplied by a conversion factor derived as follows:

(1) Since OCLC has no subject search capability, online catalog searching will be confined to known-item searches as well as searches where an author is known but not a particular title by that author. As discussed in Chapter 4 of this report, the weighted average percentage of known item searches in academic libraries is 66 percent. If an additional 10 percent is assumed to be of the known-author but unknown-item variety, a total of about 75 percent of all catalog uses are candidates for online catalog searching.

(2) Work by Tagliacozzo (10) suggests a persistence factor in catalog searches, the extent to which a user will re-access the catalog following an initial failure. The average number of catalog accesses per patron has been calculated as 1.21 searches per patron for known items and 1.52 searches per patron for unknown items.

(3) From point 1 above, it is inferred that of all searches performed online, about 13 percent will be unknown item (only author known) searches (.10

divided by .75); consequently known-item searches will be 87 percent. Applying these percentages to the persistence factor in point 2 yields 1.25 searches per applicable catalog use. Since the applicable catalog use is 75 percent of all catalog use (from point 1) the total number of searches online will be (.75 x 1.25) = 94 percent of all catalog uses reported.

(4) Given the present processing configuration at OCLC, Battelle estimated a weighted average of .78 CPU seconds per search.

(5) The CPU seconds per search multiplied by the online search factor in point 3 yields the conversion factor to derive total CPU seconds from total catalog use; in this case, (.78 x .94) = .73.

For the public libraries, a similar process takes place:

(1) The weighted average percentage of known-item searches in public libraries is 42 percent. With the additional assumption of 10 percent known-author, unknown-item searches, the candidates for online catalog searches are 52 percent of all catalog use.

(2) The persistence factor for known-item searches in public libraries is 1.21, and for unknown items it is 1.27.

(3) Calculations from the figures in (1) and (2) yield 1.22 searches per catalog use. With the applicable catalog use rate of 52 percent, the total percentage of all reported catalog searches to be performed online is (.52 x 1.22) = 63 percent.

(4) The weighted average CPU seconds per search for public libraries is .81; hence the conversion factor of CPU seconds times search percentages becomes (.81 x .63) = .51.

The total processing time to handle online catalogs for OCLC users would be about 46 CPU hours daily with the present system configuration. This is application time only; to this must be added processing overhead time required to handle and account for transactions. Assuming an overhead of 100 percent, and a 20-hour operating schedule, OCLC would require five additional CPU's to handle online public access catalogs, on the assumption of average usage rates. When the above analysis is performed for peak traffic, a total of 55 hours of applications time is needed. This additional requirement, including processing overhead, will mean six rather than five CPU's beyond OCLC's present processing requirements. Even this estimate is conservative, since peak load traffic data were not available for public libraries.

RLIN

For RLIN academic libraries, the conversion factor calculations are as follows:

(1) Since subject searching is supported at RLIN, all catalog traffic would be candidates for online catalog searching at RLIN, if such a service were to be offered. The weighted averages for known-item and unknown-item searches are 66 and 34 percent, respectively.

(2) Calculating from the persistence factors found in academic libraries yields 1.315 searches per catalog use.

(3) Figures provided to Battelle by RLIN allowed the calculation of a weighted average of .057 CPU seconds per search. The CPU seconds required per catalog use then is .057 CPU seconds per catalog search x 1.315 catalog searches per reported catalog use = .075 CPU seconds per reported catalog use.

For RLIN public libraries the same methodology is employed. The weighted search percentages for public libraries are 42 percent known-item and 58 percent unknown-item. Applying the persistence factor yields (1.21 x .42) + (1.27 x .58) = 1.245 online searches per reported catalog use. The weighted average CPU seconds per search for public libraries, given the different breakdown between known-item and unknown-item searches, is .073 CPU seconds per catalog search. Multiplying by 1.245 online searches per catalog use yields .09 CPU seconds per reported catalog use.

The figures in Figure 4A-19 show an average daily applications time of about 45 CPU minutes. Given a processing overhead time rate of 100 percent, the provision of online catalog searching by RLIN would require a full-time processor for about an hour and a half daily. Calculating peak load traffic for academic libraries raises the total time required to about 98 minutes. This is a conservative estimate for RLIN because only the academic and public libraries are taken into consideration. Special libraries, including law libraries and corporate libraries, are not included in this analysis, though they are a significant percentage of RLIN user libraries. In addition, RLIN shares a single processor with other applications at Stanford University. An additional requirement of 98 CPU minutes per day may be a significant load in this environment.

WLN

Like RLIN, WLN will allow both known-item and unknown-item searching. The weighted percentages of these searches for academic and public libraries, as well as the persistence in searching factors, are the same as for RLIN. The only difference is the weighted average CPU seconds per online catalog search. For academic libraries, this average is 0.168

CPU seconds per search; for public libraries, 0.216 CPU seconds per search. The conversion factor for academic libraries is 0.168 CPU seconds per search X 1.315 searches per catalog use = 0.22 CPU seconds per catalog use. For public libraries the factor is 0.216 X 1.245 = 0.27 CPU seconds per catalog use.

The figures in Figure 4A-20 show an average daily applications time of 48 CPU minutes for online catalog support, a figure very close to that of RLIN. Adding overhead processing time at a rate of 100 percent yields a total of 96 processing minutes per day. Given peak load traffic for academic libraries, the applications requirement becomes 53 minutes. With processing overhead of 100 percent, the total additional processing time needed is 106 minutes, or almost 2 hours of CPU time daily. Like RLIN, WLN is in a shared environment on a single processor, and an additional 2 hours may be a significant demand on resources.

References

(1) F. W. Lancaster, The Measurement and Evaluation of Library Services, New York: Information Resources Press, 1977.

(2) Ruth Hafter, "The Performance of Card Catalogs: A Review of Research, 1:3 Library Research (Fall 1979), pp. 199-222.

(3) Ben-Ami Lipetz, "Catalog Use in a Large Research Library", The Library Quarterly, 42:1 (January 1972), pp. 129-139; Richard P. Palmer, Computerizing the Card Catalog in the University Library, Littleton, Colorado, Libraries Unlimited, Inc., 1972; R. Tagliacozzo and M. Kochen, "Information-Seeding Behavior of Catalog Users", Information Storage and Retrieval, 6:5 (December 1970), pp. 363-381; Don. R.Swanson, "Requirements Study for Future Catalogs", Library Quarterly, July 1972, pp. 302-315.

(4) James. Krikelas, "Catalog Use Studies and Their Implications", in Melvin Voigt, ed., Advances in Librarianship, Volume 3, New York: Seminar Press, 1972, pp. 195-220,

(5) Brett Butler, Martha W. West and Brian Aveney, Library and Patron Response to the COM Catalog: Use and Evaluations, Los Altos, California: Information Access Corp., 1979.

(6) Helen M. Eckerd, Study of Public Libraries, LIBGIS I, Washington, D.C., National Center for Education Statistics. 1974.

(7) Walter C. Griffin, Queuing: Basic Theory and Applications, Columbus, Ohio: Grid Press, 1978.

(8) Ronald W. Force, and Jo Ellen Force, "Access to Alternative Catalogs: A Simulation Model", College and Research Libraries, 40:3 (May 1979), pp. 234-239.

(9) A. Whitney Knox, and Bruce A. Miller, "Predicting the Number of Public Computer Terminals Needed for an Online Catalog: A Queuing Theory Approach", paper presented at ALA Library Research Round Table Research Forum, Dallas, Texas, 1979.

(10) R. Tagliacozzo, L. Rosenberg, and M. Kochen, "Access and Recognition: From Users' Data to Catalog Entries", Journal of Documentation, 26 (1969), pp. 230-249.

APPENDIX 4B

WORKSHEET FOR ESTIMATING THE VOLUME OF LINKED REFERENCE SEARCHING

1. Average daily circulation			
2. Number of items circulated per use of the catalog			
3. Number of people using the catalog each day (1 ÷ 2)			
4. Number of searches by each person using the catalog			
5. Number of catalog searches each day (3x4)			
6. Percentage of searches that fail at the local catalog			
7. Number of searches that fail each day (5x6)			
8. Percentage of failed catalog searches that are searched on-line			
9. Number of failed searches that are searched on-line each day (7x8)			
10. Percentage of failed searches on the home data base			
11. Number of failed searches on the home data base each day (9x10)			
12. Percentage of failed searches that are searched on a second data base			
13. Number of failed searches that are searched on a second data base each day (11x12)			
14. Percentage of successful searches of the second data base			
15. Number of successful searches of the second data base each day (13x14)			

APPENDIX 5A

COMPUTER PROGRAMS REQUIRED FOR THE TAPE DELIVERY/NATIVE MODE LINK ALTERNATIVE

If a tape delivery/native mode link is to be established, all utilities involved would have to develop computer programs with the following functions:

Utility A

(1) Recognize user's request to search another data base

(2) Write the request statement to the offline Request File

(3) Generate Request Tape to be sent to Utility B

Utility B

(4) Process requests from the Request Tape

(5) Write search responses to offline Response File

(6) Generate Response Tape for Utility A

Utility A

(7) Write data from Response Tape into Response File

(8) Convert terminal control characters so that response may be displayed on Utility A terminals

(9) Notify user that a response is waiting*

(10) Recognize user's request to see waiting response*

(11) Transmit response to user's terminal*

Both

(12) Accounting and billing

---

* May exist now.

APPENDIX 5B

COMPUTER PROGRAMS REQUIRED FOR THE ONLINE NATIVE MODE LINK ALTERNATIVE

If an online/native mode link is to be established, all utilities involved would have to develop computer programs with the following functions:

Utility A

(1) Recognize user's request to search another data base

(2) Accept request statement and convert terminal control characters so that request can be recognized by Utility B's computer

(3) Transmit request to Utility B

Utility B

(4) Accept request, acknowledge receipt, and queue request*

(5) Transmit search response to Utility A*

Utility A

(6) Receive response

(7) Convert terminal control characters and convert from line-by-line mode to full-screen mode or vice versa so that response may be displayed on Utility A terminals

(8) Transmit response to user's terminal

Both

(9) Accounting and billing

---

* May exist now.

APPENDIX 5C

COMPUTER PROGRAMS REQUIRED FOR THE ONLINE/TRANSLATION MODE LINK ALTERNATIVE

If an online/translation mode link is to be established, all utilities involved would have to develop computer programs with the following functions:

Utility A

(1) Recognize user's request to search another data base

(2) Transmit "fill-in" screen to user's terminal

(3) Edit user's input

(4) Translate request statement into standard request format

(5) Transmit request to Utility B

Utility B

(6) Accept request and acknowledge receipt

(7) Translate request from standard format into Utility B request format, and queue request

(8) Translate search response into standard record format

(9) Transmit response to Utility A

Utility A

(10) Receive response

(11) Translate response from standard record format into Utility A record format

(12) Transmit response to user's terminal*

Both

(13) Accounting and billing

---

* May exist now.

APPENDIX 6A

SUMMARY DESCRIPTION OF BIBLINK

Introduction

BIBLINK is a computer-based tool to aid in assessing the impact of linking on each of the bibliographic utilities. A great number of "who-and-how" linking configurations can be imagined, and it is useful to compare the economic implications of many of these scenarios. BIBLINK provides a means for organizing and storing the data, speeding the calculations, and quickly generating comparisons among several linking configurations and among differing sets of assumptions.

The model is designed to be used interactively. The user "converses" with BIBLINK at a computer terminal, giving instructions that shape the direction of the session and responding to requests for data to be entered. The results of BIBLINK's calculations appear at the terminal and may also be printed offline.

The BIBLINK user begins a session by specifying a linked network configuration and by providing estimates of the interutility referral volumes that the link is expected to generate.

The model works from these basic inputs to calculate, for each utility involved, the incremental revenues and operating costs and the added burden on computer facilities.

These outputs may be produced for up to 5 years, with different data being used for each year's calculations. A different network configuration may also be specified for each of the 5 years.

BIBLINK calculations are performed by simple matrix additions and multiplications, with the appropriate rows and columns of the files being identified at the outset by the selection of a network configuration and the number of years to be simulated.

The model user can change the data quickly and easily while retaining the original data as well. "What-if" analyses may be conducted by running the model more than once and comparing the outputs.

Network Configurations

At the beginning of each model run, the user must specify the network configuration that is to be modeled. The network is "constructed" by indicating pairs of utilities and the type of link between them, selected by number from the following lists.

Pairs

1 = OCLC-RLIN
2 = OCLC-WLN
3 = OCLC-LC
4 = OCLC-5
5 = OCLC-6
6 = RLIN-WLN
7 = RLIN-LC
8 = RLIN-5
9 = RLIN-6
10 = WLN-LC
11 = WLN-5
12 = WLN-6
13 = LC-5
14 = LC-6
15 = 5-6

Link Types

1 = Tape delivery, requested items, mail
2 = Batch transmission, requested items, dial-up
3 = Online, native mode, dedicated line
4 = Online, native mode, value-added network
5 = Online, translation mode, dedicated line
6 = Online, translation mode, value-added network

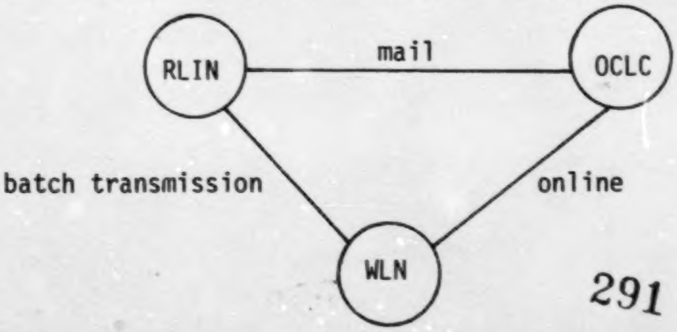
Two unnamed utilities are included: "Five" and "Six". The user may provide data for one or both, thereby adding candidates for linking that Battelle has not considered.

A network configuration may consist of any number of pairs, and each pair may employ any of the link types. Thousands of network configurations may be constructed from these elements.

Example

User selects: Pair 1 and Link 1 (OCLC-RLIN, mail tape delivery)
Pair 2 and Link 6 (OCLC-WLN, online, translation mode, via value-added network)
Pair 6 and Link 2 (RLIN-WLN, batch via dial-up line)

This configuration may be expressed schematically as follows:



BIBLINK calculates the incremental costs and revenues for each utility from each "leg" of the network specified. Thus, in the example above, the total cost to WLN of linking in this configuration is the cost associated with its RLIN link plus that associated with its OCLC link. The revenues are similarly additive.

The user may specify more than one network configuration to be modeled over the 5-year time period, and if desired, a different configuration could actually be used for each year. BIBLINK will compute the economic impact of each linking arrangement on the utilities involved. This feature permits modeling of a phased implementation of a national network over a period of several years. The user may also experiment with network evolutions that are considered likely to occur.

Input Data

The principal types of data used in the model are listed below. If BIBLINK is used to calculate only the costs of linking, and not the net revenue from linking, then the search prices may be omitted.

Search Loads

- The number of searches for cataloging, authority checking, ILL verification and location, and reference searching that would flow between each pair of utilities if a specified network configuration were implemented.

Resource Requirements

- CPU seconds required to process an average search request during an average traffic period
- Disk I/O's required to process an average search request
- Average number of characters in the search result (used only with value-added network communication method)
- Number of computer tapes sent to each other utility each day (used only with tape delivery link alternative)
- Estimated number of searches that can be entered at the user's terminal in an hour (used to calculate connect-hours of linked searching).

Operations Costs

- Average out-of-pocket cost to the utility of one CPU second
- Average out-of-pocket cost to the utility of one disk I/O
- Cost of mailing or shipping a computer tape between each pair of utilities
- Monthly cost of a dedicated leased line between each pair of utilities

- Fixed monthly and variable costs of a telephone line used for batch transmission

- Fixed monthly and variable costs of using a value-added network to establish an online link between each pair of utilities.

Search Prices

Prices charged by a utility for use of its data base by members of other utilities (some may be "0"):

- Fixed annual charge for linking

- Price per connect-hour

- Price per record transmitted to the requesting utility

- Price per record copied by someone using the requesting utility

- Price per later re-use of a record by another user of the requesting utility.

Prices charged by a utility for providing linked searching to its own users (some may be "0"):

- Price per connect-hour

- Price per request sent to a linked utility

- Price per record received and translated into the host utility's format.

Development Costs

Linking a pair of utilities will require a development process at each end. The cost of the development effort is not calculated in the model, but the user may store estimates in a Development Cost File associated with the network configuration that is being simulated. The development costs are then automatically printed out along with the incremental operations costs and revenues for each year.

All costs may be recognized in the first year of the link, or the development costs may be spread across the years of the simulation in any proportions desired. This feature permits the user to incorporate the continuing development costs of an evolving network.

Principal Outputs

The main output reports of the BIBLINK model are:

- Daily search load from linking, showing incoming and outgoing searches for each utility

- Daily resource requirements for each utility for both incoming and outgoing searches

- Annual revenues from linked searching for each utility

- Daily variable processing costs, showing the sum of data processing and transmission costs for each utility

- Statement of annual incremental revenues and costs for each utility.

Samples of these reports are shown on the following pages.

Data Sets

The data used in BIBLINK calculations are stored in files--a search load file, a search price file, a computer resources file, and so on. A simulation run uses data from each file to calculate values for the output reports.

In analyzing the impact of linking, it is useful to compare differing sets of data. The sets may incorporate differing assumptions about the search loads from linking, or they may contain data for entirely different network configurations. BIBLINK provides the capability for the user to create up to ten sets of data files, where "set" means a group of files that are used as a unit to perform one simulation run. This feature

```
             ...AVG DAILY SEARCH LOAD...

                 OUTGOING, YEAR1

          TOTAL      CAT      AUT      ILL      REF

OCLC          0        0        0        0        0
RLIN          0        0        0        0        0
WLN           0        0        0        0        0
LC            0        0        0        0        0

                 INCOMING, YEAR1

          TOTAL      CAT      AUT      ILL      REF

OCLC          0        0        0        0        0
RLIN          0        0        0        0        0
WLN           0        0        0        0        0
LC            0        0        0        0        0




       ...ANNUAL INCOMING SEARCH REVENUES...

                      YEAR1

          TOTAL      CAT      AUT      ILL      REF

OCLC          0        0        0        0        0
RLIN          0        0        0        0        0
WLN           0        0        0        0        0
LC            0        0        0        0        0
```

```
...VAR. DAILY PROCESSING COSTS...

            TOTAL, YEAR1

                  AVGE

  OCLC            0.00
  RLIN            0.00
  WLN             0.00
  LC              0.00

           OUTGOING, YEAR1

                  AVGE

  OCLC            0.00
  RLIN            0.00
  WLN             0.00
  LC              0.00

           INCOMING, YEAR1

                  AVGE

  OCLC            0.00
  RLIN            0.00
  WLN             0.00
  LC              0.00
```

```
...AVG DAILY OP REQUIREMENTS..

   OCLC, INCOMING, YEAR1

             CPU*SECS  DISK*IO

 TOTAL              0          0
 CAT                0          0
 AUT                0          0
 ILL                0          0
 REF                0          0

   RLIN, INCOMING, YEAR1

             CPU*SECS  DISK*IO

 TOTAL              0          0
 CAT                0          0
 AUT                0          0
 ILL                0          0
 REF                0          0

   WLN, INCOMING, YEAR1

             CPU*SECS  DISK*IO

 TOTAL              0          0
 CAT                0          0
 AUT                0          0
 ILL                0          0
 REF                0          0

   LC, INCOMING, YEAR1

             CPU*SECS  DISK*IO

 TOTAL              0          0
 CAT                0          0
 AUT                0          0
 ILL                0          0
 REF                0          0
```

```
                  ...AVG DAILY OP REQUIREMENTS...

                      OCLC, OUTGOING, YEAR1

              CPU*SECS  DISK*IO AVG*CHAR    TAPES PORT*HRS

TOTAL               0         0         0        0        0
CAT                 0         0         0        0        0
AUT                 0         0         0        0        0
ILL                 0         0         0        0        0
REF                 0         0         0        0        0

                      RLIN, OUTGOING, YEAR1

              CPU*SECS  DISK*IO AVG*CHAR    TAPES PORT*HRS

TOTAL               0         0         0        0        0
CAT                 0         0         0        0        0
AUT                 0         0         0        0        0
ILL                 0         0         0        0        0
REF                 0         0         0        0        0

                      WLN, OUTGOING, YEAR1

              CPU*SECS  DISK*IO AVG*CHAR    TAPES PORT*HRS

TOTAL               0         0         0        0        0
CAT                 0         0         0        0        0
AUT                 0         0         0        0        0
ILL                 0         0         0        0        0
REF                 0         0         0        0        0

                      LC, OUTGOING, YEAR1

              CPU*SECS  DISK*IO AVG*CHAR    TAPES PORT*HRS

TOTAL               0         0         0        0        0
CAT                 0         0         0        0        0
AUT                 0         0         0        0        0
ILL                 0         0         0        0        0
REF                 0         0         0        0        0
```

```
                 STATEMENT OF INCREMENTAL REVENUES AND COSTS

                                 RLIN

                     YEAR1     YEAR2     YEAR3     YEAR4     YEAR5     TOTAL

                              HOME REVENUES

CAT                      0         0         0         0         0         0
AUT                      0         0         0         0         0         0
ILL                      0         0         0         0         0         0
REF                      0         0         0         0         0         0


                              LINK REVENUES

CAT                      0         0         0         0         0         0
AUT                      0         0         0         0         0         0
ILL                      0         0         0         0         0         0
REF                      0         0         0         0         0         0
LINK FEES                0         0         0         0         0         0

TOTAL REVENUES           0         0         0         0         0         0


                             OPERATING COSTS

OP*INCOMING              0         0         0         0         0         0
OP*OUTGOING              0         0         0         0         0         0
DATA*TRANS               0         0         0         0         0         0
OTHER*COSTS              0         0         0         0         0         0

TOTAL OP COSTS           0         0         0         0         0         0


                          NET REVENUES OVER COSTS

NET REVENUES             0         0         0         0         0         0


                            DEVELOPMENT COSTS

PLAN+DSGN                0         0         0         0         0         0
HARDWARE                 0         0         0         0         0         0
SOFTWARE                 0         0         0         0         0         0
SITE*PREP                0         0         0         0         0         0
INSTALL                  0         0         0         0         0         0
TESTING                  0         0         0         0         0         0
ACCOUNTNG                0         0         0         0         0         0
OTHER                    0         0         0         0         0         0

TOTAL DV COSTS           0         0         0         0         0         0
```

allows as many as ten separate linking scenarios to be stored in BIBLINK's data base at one time.

BIBLINK can also be used to perform "what-if" analyses. For example, a model user might be interested in seeing the effect on a utility's net revenues if the incoming search load were to be higher by 1,000 than the value in the search load file. It is not necessary to create a whole new set of files to incorporate this simple adjustment. The user may change the data in question temporarily, run the model again immediately, and compare the "before" and "after" results.

APPENDIX 6B

LINKED SEARCH LOAD ESTIMATES

This appendix contains the search load estimates used in assessing the economic impact of linking the utilities. A computer-based simulation model, described in Appendix 6A, was used to project the computer resource requirements and operations costs resulting from these estimated search loads.

Online/Translation Link Alternative

Figures 6B-1 through 6B-4 present the search load estimates for the online translation link alternative. The linked search loads for shared cataloging, ILL, reference searches, and authority checks are estimated for each of the utilities. The derivation of these volumes did not attempt to account for the effects of charges for linked searching. It was assumed that these charges will be reasonable and will not be a major determinant in the decision to use or not to use the linked resources.

Online/Native Mode Link Alternatives

It has been assumed that the difficulty of using a native mode link would dampen the demand for linked searching. The search load estimates for this link alternative were calculated as 90 percent of the loads estimated for the online/translation link.

Tape Delivery Link Alternative

Cataloging search load is assumed to be the same as that estimated for the online/native mode alternative (90 percent of the search loads in Figure 6B-1). The delay of 1 to 2 weeks may be acceptable to many catalogers.

Only token volume is assumed for authority searches, reference searches, and ILL verification and location searches. The delay would make linked searching much less desirable than online searching for these purposes. The following search loads were used:

Linked Searches Per Day, All Libraries

	Authority	ILL	Reference	Total
OCLC to RLIN	30	100	100	230
RLIN to OCLC	0	10	10	20
OCLC to WLN	15	5	10	30
WLN to OCLC	0	5	5	10
OCLC to LC	30	0	0	30
LC to OCLC	0	0	0	0
RLIN to WLN	30	0	50	80
WLN to RLIN	10	0	5	15
RLIN to LC	70	0	0	70
LC to RLIN	0	0	0	0
WLN to LC	10	0	0	10
LC to WLN	0	0	0	0
TOTAL	195	120	180	495

Since these are only nominal volumes, 1980 estimates for authority, ILL, and reference searching were also used for 1983.

FIGURE 6B-1. ESTIMATED NUMBER OF LINKED SEARCHES FOR CATALOGING PER DAY--ONLINE/TRANSLATION ALTERNATIVE

OCLC Cataloging		1980	1983	Assumptions
Original records entered per day		3,200	5,800	1980: about 1 million new records added, divided by 313 days 1983: about 1.8 million new records added, divided by 313 days
Percent of titles that would be searched on other data bases		25	25	OCLC users would add only 1 percentage point to current title hit rate on each data base searched. When this low probability of finding a record becomes generally known, the users will not search many of the titles on other data bases, knowing that most would not be found. Linked searching may stabilize at around 25 percent of titles.
Linked searches per day		800	1,450	
Number of linked searches on each data base	OCLC			
	RLIN	550	1,000	Most search RLIN first because of larger data base than WLN
	WLN	50	100	Regional material; titles not hit on RLIN.
	LC	200	350	For preliminary LC-MARC records
RLIN Cataloging		1980	1983	Assumptions
Original records entered per day		1,200	1,450	1980: about 375,000 new records added, divided by 313 days 1983: RLG estimates about 450,000 new records added, divided by 313 days
Percent of titles that would be searched on other data bases		50	33	1980: Although many no-hits can be found on other data bases, especially OCLC, research libraries may choose to do original cataloging for much material. 1983: RLIN hit rate is higher than 1980, and fewer titles need to be searched on other data bases.
Linked searches per day		500	480	
Number of linked searches on each data base	OCLC	500	400	Most search OCLC first because of large data base
	RLIN			
	WLN	50	40	Regional material; titles not hit on OCLC
	LC	50	40	For preliminary LC-MARC records
WLN Cataloging		1980	1983	Assumptions
Original records entered per day		270	540	1980: WLN estimates about 85,000 new records added, divided by 313 1983: WLN estimates about 170,000 new records added, divided by 313
Percent of titles that would be searched on other data bases		90	90	At least half of current titles that don't hit on WLN may be found on OCLC
Linked searches per day		240	480	
Number of searches on each data base	OCLC	200	400	Most search OCLC first because of data base size and similarity of membership
	RLIN	20	40	OCLC no-hits; first searches by research libraries
	WLN			
	LC	20	40	For preliminary LC-MARC records
LC Cataloging		1980	1983	Assumptions
Linked searches per day		500	600	Most Roman alphabet titles are searched
Number of searches on each data base	OCLC	500	575	OCLC is searched first because of large data base, many libs.
	RLIN		25	As RLIN data base grows, some titles are searched there also
	WLN			
	LC			

FIGURE 6B-2. ESTIMATED NUMBER OF LINKED SEARCHES FOR ILL PER DAY--ONLINE/TRANSLATION ALTERNATIVE

OCLC ILL		1980	1983	Assumptions
Monograph borrowing requests searched online each day		900	1,250	1980: Average 200 requests a year(1) x 1,430 libraries(2), divided by 313 days. 1983 Projection from 1980 figure, assuming expanded membership
Percent of titles that would be searched on other data bases		25	25	Chapter 3 shows that only 23 percent of requests find no holdings on OCLC. See Figure 3-4.
Linked searches per day		225	315	
Number of searches on each data base	OCLC			
	RLIN	200	265	
	WLN	25	50	
	LC			

RLIN ILL		1980	1983	Assumptions
Monograph borrowing requests searched online each day		100	150	1980: Average 500 requests a year(1) x 65 libraries(2), divided by 313 days 1983: Projection from 1980 figures, assuming expanded membership
Percent of titles that would be searched on other data bases		70	50	1980: Chapter 3 shows that about 70 percent of requests find no holdings on RLIN. See Figure 3-4. 1983: Expected growth of RLIN data base should reduce percentage of no-hits
Linked searches per day		70	75	
	OCLC	60	65	Most search OCLC first because of great number of holdings
	RLIN			
	WLN	10	10	
	LC			

WLN ILL		1980	1983	Assumptions
Monograph borrowing requests searched online each day		40	50	1980: Average 200 requests a year(1) x 59 libraries(2), divided by 313 days 1983: Projection from 1980 figure, assuming expanded membership
Percent of titles that would be searched on other data bases		70	50	1980: Chapter 3 shows that about 70 percent of requests find no holdings on WLN. See Figure 3-4. 1983: Expected growth of WLN data base should reduce percentage of no-hits.
Linked searches per day		28	25	
Number of searches on each data base	OCLC	25	20	Most search OCLC first because of greater number of holdings.
	RLIN	3	5	
	WLN			
	LC			

LC ILL		1980	1983	Assumptions
Linked searches per day		10	10	Battelle estimate
Number of searches on each data base	OCLC	8	8	
	RLIN	1	1	
	WLN	1	1	
	LC			

(1) Estimates used in Chapter 3. See Figures 3-9 and 3-10.
(2) Public, academic, and research libraries only, mid-1980.

FIGURE 6B-3. ESTIMATED NUMBER OF LINKED REFERENCE SEARCHES* PER DAY--ONLINE/TRANSLATION ALTERNATIVE

OCLC Reference		1980	1983	Assumptions
Potential number of linked reference searches each day		33,700	38,000	1980: Daily demand for linked reference searching from Chapter 4. See Figure 4-2. 1983: Projection based on 1980 figure.
Percent that would be searched on other data bases		50	50	Hit rates are likely to be quite low, and demand for linked reference searching may stabilize at about 50 percent when experience shows that few searches are successful.
Linked searches per day		16,850	19,000	
Number of searches on each data base	OCLC			
	RLIN	14,850	17,000	Most search RLIN first, since most linked searching is done by research libraries
	WLN	2,000	2,000	
	LC			

RLIN Reference		1980	1983	Assumptions
Potential number of linked reference searches each day		3,600	4,500	1980: Daily demand for linked reference searching from Chapter 4. See Figure 4-2. 1983: Projection based on 1980 figure.
Percent that would be searched on other data bases		50	50	Hit rates are low in 1980, and demand for linked reference searching may stabilize at 50 percent when experience shows that few searches are successful. Growth of RLIN data base by 1983 should slow the increase in demand for linked reference searching; half will still search another data base.
Linked searches per day		1,800	2,250	
Number of searches on each data base	OCLC	800	850	OCLC searched first except for subject searches
	RLIN			
	WLN	1,000	1,400	All subject searches
	LC			

WLN Reference		1980	1983	Assumptions
Potential number of linked reference searches each day		650	800	1980: Daily demand for linked reference searching from Chapter 4. See Figure 4-2. 1983: Projection based on 1980 figure.
Percent that would be searched on other data bases		50	50	Hit rates are likely to be quite low and demand for linked reference searching may stabilize at about 50 percent when experience shows that few searches are successful.
Linked searches per day		325	400	
Number of searches on each data base	OCLC	150	200	OCLC searched first except for subject searches
	RLIN	175	200	All subject searches
	WLN			
	LC			

LC Reference		1980	1983	Assumptions
Linked searches per day		10	10	Battelle estimate
	OCLC			
	RLIN	10	10	
	WLN			
	LC			

* Includes only those online searches that result from failure to find anything in the local catalog. Actual level of linked reference searching will be higher.

FIGURE 6B-4. ESTIMATED NUMBER OF LINKED AUTHORITY FILE SEARCHES PER DAY

	1980	1983
OCLC to RLIN	9,600	12,000
RLIN to OCLC	0	0
OCLC to WLN	2,400	3,000
WLN to OCLC	0	0
OCLC to LC	0	0
LC to OCLC	0	0
RLIN to WLN	800	1,250
WLN to RLIN	300	375
RLIN to LC	0	0
LC to RLIN	0	0
WLN to LC	0	0
LC to WLN	0	0
TOTAL	13,100	16,625

APPENDIX 6C

OVERVIEW OF THE BIBLIOGRAPHIC UTILITIES

(All information provided by the organizations in July 1980)

CHARACTERISTIC	OCLC	RLG/RLIN	WLN	LC
Services				
• Shared Cataloging	Yes	Yes	Yes	Yes
• Interlibrary Loan	Yes	(Early 1981)	(1981)	No
• Subject Searching	No	Yes	Yes	Yes
• Authority	LC name	Partial	Yes	Yes
• Acquisitions	(Fall 1980)	Yes	Yes	Yes
• Serials	Yes	Yes	Partial	No
• Circulation	(Under development)	No	Yes	No
• Patron Access	Yes	Stanford University	Wash. State Library	No
Network Scope				
• Service Area	U.S. and Canada	U.S.	Northwestern U.S.	Internal
• Affiliated Libraries	2181 (7/80)	RLG: 20 (7/80) RLIN: 207 (7/80)	59 (6/80)	N/A
• Regional Networks	20	N/A	2	N/A
• Terminals (dedicated, dial-up)	3290, 528	195, 150	142, 7	1,126
• Lines	165	15	20	191
Hardware and Software				
• Host Computer(s)				
- Type	Sigma 9 (5), Tandem (14)	IBM 3033	Amdahl V6-II	Amdahl 470, IBM 3033
- Percent Dedicated	100%	20%	23%	100%
- Memory Size	Sigma: 256K words Tandem: 384 million bytes	6 megabytes	10 megabytes	Amdahl: 6 megabytes IBM: 8 megabytes
- Disks	30 mirrored 300 meg	9 3350's	9 3350's	8,349 million bytes
- Software	Modified RBM, CATROOT, MASTIO/T/TOS, GUARDIAN, ENVOY, ENSCRIBE	SPIRES, MILTEN, WYLBUR, ORVYL, BALLOTS, PL360	ADABAS, PL/I, CICS, MVS	SCORPIO/MUMS, CICS
• Front-End Processor	D-116 (16), Tandem (5)	PDP 11/60	COMTEN	COMTEN 3670 (2)
• Network				
- Line Type	Dedicated, polled, multidrop	Leased, polled, multi-drop, synchronous	Leased, polled, multi-drop, synchronous	Leased, polled, multi-drop, synchronous
- Transmission Rate	2400 BPS	2400 BPS	4800 BPS	2400 BPS
• Terminals	OCLC 100 and OCLC 105 (Beehives), OCLC 200 (Ramtek, 9/80)	Zentec 9003, Zentec ZMS-90	Hazeltine	Four-Phase Sycor
Data Base				
• Type	Hashed with linked lists in index	Integrated tree structure (SPIRES)	Integrated, linked (ADABAS)	Not provided
• Files	Bibliographic Record, Institution Holding, Local Data Record, Authority, Interlibrary Loan Record	Book System Files; LC-MARC (1972 on), Catalog Data, Stanford In Process Non-Book System Files: LC-MARC (films, maps, serials), CONSER snapshot, Non-Book Catalog Data	Bibliographic Files: Bibliographic, Holdings, Authority, Working, Waiting Acquisitions Files: In-Process, Standing Orders, Name and Address, Account States	LC-MARC (books, serials, maps), In-Process, Name Authority
• Number of Records (7/80)				
- Bibliographic	6.4 million, 3000/day	1.6 million book 141,000 non-book	1.7 million, 15,000/mo.	1 million, 200K/year
- Holdings	72.7 million	1.8 million	1.5 million	N/A
- Authority	281,430	not provided	2.1 million	35,000 names
- Serials	350,000	140,000	N/A	N/A
• Transaction Volume (7/80)	1.6 million per day	80,000 per month	37,000 per day	1.9 million per month
• Average Response Time	6.5 seconds	2-3 seconds	2-4 seconds	4.4 seconds

APPENDIX 6D

IMPLEMENTATION PLAN FOR THE TAPE DELIVERY/NATIVE MODE LINK ALTERNATIVE

Assumptions

(a) Search requests and responses are transferred in the format of the receiving utility (in that utility's "native mode").

(b) The host utility's computer manipulates the message format so that the linked utility's computer recognizes it as a request. Response formats are also manipulated so that the responses may be displayed on the host utility's terminals.

Tasks	Development Time per Utility
1. Develop Standards	
1.1 Appoint an Applications Standards Committee	
1.2 Meetings of the Applications Standards Committee	3 person-months
1.2.1 Review message header formats used by each utility	
1.2.2 Review request statement formats used by each utility	
1.2.3 Determine manipulation required to enable requests entered on a host utility terminal to be recognized by the linked utility's computer as incoming requests	
1.2.4 Determine manipulation required to enable responses from the linked utility to be displayed on host utility terminals; develop terminal control character conversion tables; examine requirements for converting from line-by-line mode to screen-by-screen mode, or vice versa	
1.2.5 Determine accounting, charging, and billing arrangements	

APPENDIX 6D (Continued)

IMPLEMENTATION PLAN FOR THE TAPE DELIVERY/NATIVE MODE LINK ALTERNATIVE

	Tasks	Development Time per Utility
2.	Design, Develop, and Test Applications Software	
2.1	Program(s) to recognize user's request to search another data base, to manipulate the format, and to write the request statement to the offline Request File	6 person-months
2.2	Program to generate Request Tapes to be sent to other utilities	1 person-month
2.3	Program(s) to process Request Tapes received from other utilities, and to write search responses to offline Response File	2 person-months
2.4	Program to generate Response Tape to be returned to each requesting utility	1 person-month
2.5	Program(s) to write data from Response Tapes into Response File at the host utility, notify user of waiting response	4 person-months
2.6	Program to manipulate response message so that it may be displayed on host utility's terminals	8 person-months
2.7	Accounting and billing program(s)	3 person-months
	TOTAL	28 person-months per utility 2.3 person-years per utility

APPENDIX 6E

IMPLEMENTATION PLAN FOR THE ONLINE/NATIVE MODE LINK ALTERNATIVE

Assumptions

(a) Search requests and responses are transferred in the format of the receiving utility (in that utility's "native mode").

(b) The host utility's computer manipulates the message format so that the linked utility's computer recognizes it as a request. Response formats are also manipulated so that the responses may be displayed on the host utility's terminals.

(c) Communications interface software is developed independently at each utility.

Tasks	Development Time per Utility
1. Develop Standards	
1.1 Appoint an Applications Standards Committee and a Communications Standards Committee	
1.2 Meetings of the Applications Standards Committee	3 person-months
1.2.1 Review message header formats used by each utility	
1.2.2 Review request statement formats used by each utility	
1.2.3 Determine manipulation required to enable requests entered on a host utility terminal to be recognized by the linked utility's computer as incoming requests	
1.2.4 Determine manipulation required to enable responses from the linked utility to be displayed on host utility terminals; develop terminal control character conversion tables; examine requirements for converting from line-by-line mode to screen-by-screen mode, or vice versa	
1.2.5 Determine accounting, charging, and billing arrangements	

APPENDIX 6E (Continued)

IMPLEMENTATION PLAN FOR THE ONLINE/NATIVE MODE LINK ALTERNATIVE

Tasks	Development Time per Utility
1.3 Meetings of the Communications Standards Committee	12 person-months
1.3.1 Determine the suitability of the NCLIS/NBS protocol for use in an online/native mode link	
1.3.2 Define an applications level protocol, perhaps a subset of the full NCLIS/NBS protocol; determine a minimum protocol to be used in a prototype test	
1.3.3 Evaluate communications hardware and software; select lower level protocols (e.g., bisynchronous interface between TANDEM and COMTEN communications processors); select communications hardware)	
2. Design, Develop, and Test Applications Software	
2.1 Program to recognize user's request to search another data base	6 person-months
2.2 Program(s) to accept request statement and manipulate the format so that it can be recognized by the linked utility's computer as an incoming request	6 person-months
2.3 Program(s) to receive response from linked utility, manipulate message format so that it may be displayed on host utility terminals, and transmit response to user's terminal	12 person-months
2.4 Accounting and billing program(s)	3 person months
3. Design, Develop, and Test Communications Software	24-48 person months
3.1 Develop software specifications for the applications level protocol	
3.2 Program the protocols	

APPENDIX 6E (Continued)

IMPLEMENTATION PLAN FOR THE ONLINE/NATIVE MODE LINK ALTERNATIVE

Tasks	Development Time per Utility
3.3 Install and test communications hardware	
3.4 Test the communications link	
TOTAL	66-90 person-months per utility 5.5-7.5 person-years per utility